DOCUMENT RESUME

ED 216 426 EA 014 596

AUTHOR Capell, Frank J.
TITLE A Study of Alternatives in American Education, Vol. VI: Student Outcomes at Alum Rock 1974-1976.
INSTITUTION Rand Corp., Santa Monica, Calif.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
REPORT NO ISBN-0-8330-0042-X (v. 1); Rand-R-2170/6-NIE
PUB DATE Jul 81
CONTRACT NIE-B2C-5326
NOTE 121p.; For related documents, see ED 163 591, ED 166 825, ED 206 058-059, EA 014 595 and EA 014 597.
AVAILABLE FROM Publications Department, The Rand Corporation, 1700 Main Street, Santa Monica, CA 90406 ($10.00).

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Comparative Analysis; Educational Change; Educational Vouchers; Elementary Education; *Nontraditional Education; *Outcomes of Education; Participative Decision Making; Public Schools; *School Choice; School Size; Tables (Data); Teacher Attitudes
IDENTIFIERS *Alum Rock Union School District CA

ABSTRACT

If interpretations of data compiled during a five-year demonstration of alternative educational programing in the elementary schools of California's Alum Rock School District are correct, the nature of, the size of, and the parental choice exercized over alternative programs have no significant effect on students' reading achievement, social skills, or perceptions of themselves and others. Originally an experimental educational voucher system within the public school system, the Alum Rock project never became a full-fledged voucher program but did enable researchers to examine the relative cognitive and noncognitive impacts of several variables in educational programing. Neither the experimental group of students (those in alternative programs) nor the control group (those remaining in regular programs) was a random sample, information was lacking concerning the motivations behind parental choices, and test data were imperfect, so researchers had to modify their information to accommodate sample biases before applying analysis of covariance. Over 9,000 students were involved in the study, 60 percent of whom enrolled in alternative programs. The results, presented with caution, indicate little difference in the benefits to students of alternative and regular programs, but show a marked relationship between student outcomes and teachers' perceptions of themselves and their programs. (Author/PGD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED216426

A STUDY OF ALTERNATIVES IN AMERICAN EDUCATION, VOL. VI: STUDENT OUTCOMES AT ALUM ROCK 1974-1976

PREPARED FOR THE NATIONAL INSTITUTE OF EDUCATION

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Tom Cockrell

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FRANK J. CAPELL

WITH THE ASSISTANCE OF LYNN DOSCHER

R-2170/6-NIE
JULY 1981

EA 014 596

Rand
SANTA MONICA, CA 90406

The work upon which this publication is based was performed pursuant to Contract B2C-5326 with the National Institute of Education, Department of Health, Education, and Welfare.

Library of Congress Cataloging in Publication Data (Revised)

Main entry under title:

A Study of alternatives in American education.

"R-2170/
Includes bibliographies.
Contents: v. 1. Bass, G. V. District policies and the implementation of change. -- -- v. 3. Rasmussen, R. Teachers' responses to alternatives.--[etc.] -- v. 6. Capell, F.J., Doscher, L. Student outcomes at Alum Rock, 1974-1976.
1. Educational innovations--United States--Case studies.
2. Non-formal education--United States--Case studies.
3. School districts--United States--Case studies.

I. National Institute of Education. II. Rand Corporation.
III. Title: Alternatives in American education.
AS36.R3 R-2170/1, etc. [LA217] 081s 78-14761
ISBN 0-8330-0042-X (v. 1) [379.1'535'0973] AACR2

The Rand Publications Series: The Report is the principal publication documenting and transmitting Rand's major research findings and final research results. The Rand Note reports other outputs of sponsored research for general distribution. Publications of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.

Published by The Rand Corporation

A STUDY OF ALTERNATIVES IN AMERICAN EDUCATION, VOL. VI: STUDENT OUTCOMES AT ALUM ROCK 1974-1976

PREPARED FOR THE NATIONAL INSTITUTE OF EDUCATION

FRANK J. CAPELL

WITH THE ASSISTANCE OF LYNN DOSCHER

R-2170/6-NIE
JULY 1981

Rand
SANTA MONICA, CA. 90406

10.00

PREFACE

This report is the sixth volume of a series documenting a study of alternative schools in American education, sponsored by the National Institute of Education under Contract B2C-5326. There are six other volumes in the series, all published or forthcoming under the general title, A Study of Alternatives in American Education:

Vol. I: District Policies and the Implementation of Change, by G. Bass, R-2170/1-NIE

Vol. II: The Role of the Principal, by Margaret A. Thomas, R-2170/2-NIE

Vol. III: Teachers' Responses to Alternatives, by R. Rasmussen, R-2170/3-NIE

Vol. IV: Family Choice in Schooling, by R. G. Bridge and J. Blackman, R-2170/4-NIE

Vol. V: Diversity in the Classroom, by P. Barker, T. K. Bikson, and J. Kimbrough, R-2170/5-NIE

Vol. VII: Summary and Policy Implications, by the Educational and Human Resources Program, R-2170/7-NIE

This study has its origins in 1972. In April of that year, the Office of Economic Opportunity (OEO) funded an education voucher demonstration in Alum Rock, California, and awarded a study and evaluation contract to The Rand Corporation.[1] Voucher systems require that funds for education be distributed directly to families in the form

[1] Findings for the first year of the voucher demonstration (1972-73) are reported in Daniel Weiler et al., A Public School Voucher Demonstration: The First Year at Alum Rock, The Rand Corporation, R-1495-NIE, June 1974, 4 vols. Alum Rock is an independent elementary school district in San Jose, California.

of certificates, which families can then use to purchase education at schools of their choice. The government wished to test a voucher model that included competing public and private schools, with complex regulations designed to protect and advance the interests of disadvantaged families.[2] But the OEO agreement with Alum Rock did not require immediate implementation of this model. In lieu of private schools participating in the demonstration, Alum Rock was to encourage parent choice and stimulate competition between schools--two key objectives of the voucher plan--by creating multiple programs within the public schools. Parents would be informed about their options and encouraged to select the programs they preferred for their children. Alum Rock and OEO agreed that this "public schools only" model was to be a "transition" toward a more complete voucher demonstration, and OEO continued to seek additional demonstration sites for a more extensive test of the voucher idea. The demonstration began in September 1971 with six schools, organized as twenty-two "minischools" offering a variety of educational approaches.

By the end of the second year of the demonstration--spring 1974--sponsorship of the voucher program had been assumed by the National Institute of Education. The transition to a full-scale model in Alum Rock had not taken place, and no new sites had joined the demonstration. Rand and NIE agreed, however, that while a more complete voucher test

---

[2] The "regulated compensatory" voucher model was originally proposed in a 1970 study commissioned by OEO. See Center for the Study of Public Policy, Education Vouchers: A Report on Financing Elementary Education by Grants to Parents, Cambridge, Mass., December 1970.

might still be arranged in Alum Rock or elsewhere, the existing demonstration was of interest in its own right: Thirteen public schools were offering forty-five program options to parents.[3] In effect, Alum Rock was testing a variant of an innovation that a number of observers had argued could improve the quality of public education--alternative schools.

It was agreed that while the main study would continue to concentrate on Alum Rock in 1974-75, a small side study would be undertaken to explore the nature of the alternative schools movement in other districts. This study identified a number of areas where further analysis might yield a better understanding of the issues associated with implementing alternative schools. Many of these issues had already surfaced in Alum Rock.

By the fourth year of the demonstration (1975-76), prospects for creating a more comprehensive test of the voucher model had diminished appreciably, while the work that had already been accomplished in Alum Rock constituted a useful base for a modest comparative study of alternative schools. Accordingly, some project resources were shifted in that year toward the study of three new sites where alternative schools were being tried: Cincinnati, Ohio; Eugene, Oregon; and Minneapolis, Minnesota.[4] Data collection from these sites and Alum Rock was completed in 1976-77.

---

[3] There were at one time more than fifty minischools available to participating parents, in fourteen demonstration schools. Ten Alum Rock schools never joined the demonstration.

[4] Criteria and methods for site selection are discussed in Chapter 1 of Vol. 1 in this series: District Policies and the Implementation of Change, by G. Bass.

In this report, we attempt to provide answers to two broad questions regarding student cognitive and noncognitive outcomes of the Alum Rock demonstration. One question asks whether the demonstration affected student outcomes differently in alternative schools and regular schools. The second question focuses on alternative schools in Alum Rock and asks whether there were particular features of the educational processes in these schools that were associated with variation in student outcomes--more specifically, "Were perceptions and attitudes of teachers, characteristics of alternative-school programs, perceptions of students, and choice of programs by parents associated with differences in student outcomes?"

SUMMARY

During the last two decades, school districts across the country have responded to pressures for parental choice and educational improvement by instituting various kinds of educational alternatives. In California, one such district was the Alum Rock United School district. Alum Rock implemented a wide variety of alternative-school programs as part of a voucher demonstration funded by the Office of Economic Opportunity (and later the National Institute of Education). The Rand Corporation was awarded a contract to study and evaluate the demonstration.

Although Alum Rock did not, finally, implement a true voucher demonstration, Rand was able to evaluate the effects of various school reforms on students' cognitive and affective outcomes. These reforms included parental choice, alternative-education programs (in the form of mini-schools), smaller-sized schools, and decentralized decisionmaking (which increased teachers' and principals' authority over budgets and curricula).

The study attempted to answer two broad questions: Were students' cognitive and noncognitive outcomes different in alternative and regular schools? In alternative schools, were cognitive outcomes affected by parental choices, program size, students' perceptions of their classroom environment, and teachers' perceptions of alternatives and attitudes toward them? We used students' reading test scores as the cognitive outcome and their social, self, and peer perceptions as the noncognitive outcomes. In our analyses of these data, we controlled for the

following student-background variables: the relevant cognitive or noncognitive pretest, ethnicity, socioeconomic status, language spoken at home, and number of days in school.

The demonstration presented problems that affected our analyses and interpretation of student outcomes. Data on students in regular schools for the first two years were, for several reasons, unusable. Students and classes could not be randomly assigned to regular and alternative schools. Thus data had to be adjusted statistically for differences in students enrolled in each, making it necessary to interpret the effects cautiously. Further, the alternative programs had been in operation for, at most, three years; thus, estimates of their effect on outcomes are necessarily incomplete. These limitations should be kept in mind.

The conclusions and implications summarized here should help policymakers decide how desirable parental choice, alternative school programs, etc. are for implementing school improvements. They should also help other school districts implement and improve education. However, we caution that cognitive and noncognitive student outcomes were simply potential (though desirable) by-products of the Alum Rock demonstration. They should not be used as the sole basis for evaluating attributes of school reform or delivery systems such as vouchers or tax credits.

Do student outcomes differ in alternative and regular schools? We found no appreciable or consistent differences in students' (adjusted) reading achievement between regular and alternative schools. The same was true for two of our noncognitive outcomes--self-esteem and perception of peers. However, students in alternative schools perceived

themselves as very slightly more distant from significant others (such as teachers) than did students in regular schools (a finding of statistical rather than practical or policy significance).

Our second question asked how certain alternative-school features, such as parental choice, program size, and students' and teachers' perceptions and attitudes, affected students' cognitive outcomes.

To test the assumption that parental choice would indirectly affect achievement by making a better match between students' needs and the education they receive, we examined the effect of that choice on reading achievement. Because we had no information to indicate how actively or on what basis parents made a particular placement, we used proxies for those data (students' number of program changes and nonlocal school attendance). As measured by those proxy variables, parental choice appears unrelated to student achievement.

A second feature of alternative schools is reduced class size. We found that smaller program size had a slightly positive effect on reading achievement in the 1975-1976 school year, but school size was unrelated to achievement in the previous school year.

The relationship between students' perceptions of their program and their achievement was mixed. As students' perceptions of the difficulty of classwork increased, mean scores on reading achievement increased slightly. However, their perceptions of the social environment and organization of their class had no appreciable or consistent effects on reading scores.

The effect of teachers' perceptions on reading achievement proved to be the most complex and significant feature we studied. We examined

that effect on two levels--the minischool and the classroom. On the surface, the results appear contradictory, but the apparent contradiction results from the different focus at each level.

At the minischool level, teachers' perceptions reflect the implementation of the alternative program, particularly decentralized decisionmaking. We found that minischools whose teachers, on average, perceived that their program was cohesive, directed by common policy, and actively involved the principal had higher reading achievement than minischools whose teachers perceived the opposite. Minischools whose teachers, on average, saw themselves as having greater autonomy and influence at the program level were associated with lower reading achievement.

At the classroom level, teachers' perceptions reflected the individual differences among teachers within the minischools. Those classrooms within minischools whose teachers saw themselves as more autonomous and influential than their colleagues had higher reading achievement, on average, regardless of minischool program or the mean reading achievement for the program as a whole. In other words, regardless of whether these teachers taught in cohesive or noncohesive minischools, their classes would be likely to have higher reading achievement.

Although limited in scope, this study has instructive implications for those interested in educational alternatives. First, experimenting with parental choice and the nature and size of programs had no apparent effect on students' reading achievement, perceptions of themselves and others, or social skills. Thus, debate over educational alternatives

should be based on community interests or public policy rather than the possible effects on student outcomes.

Second, teachers' perceptions evidently do affect student reading achievement. Thus, program implementation and teacher selection should take these perceptions into account. Program implementation should lead to cohesiveness, shared policy, and principal's support. However, in selecting teachers, districts should pay attention to possible tradeoffs between autonomy and cohesiveness. Our results indicate that the classrooms of autonomous, influential teachers have higher reading achievement, but that it may come at the expense of a program's overall effectiveness.

ACKNOWLEDGMENTS

The advisory panel for this project--Stephen Klein, Robert L. Linn, Richard J. Shavelson, Rodney Skager, and Daniel Weiler--gave valuable guidance and support. Special thanks for continued advice and encouragement throughout the work go to Leigh Burstein, Robert L. Linn, Roger Rasmussen, Richard L. Shavelson, and Dan Weiler.

CONTENTS

PREFACE iii

SUMMARY vii

ACKNOWLEDGMENTS xiii

TABLES xvii

FIGURES xix

Section

I. INTRODUCTION 1
Role of Student Outcomes 2
Features of the Alum Rock Demonstration and Their Relation to Student Outcomes 4
Overview of the Research 6
Scope and Limitations of the Research 12
Outline of the Report 13

II. METHODS 14
Description of the Sample 14
Dependent Variables 18
Covariates and Independent Variables 20
Student Performance on the Reading Tests 25
Data Analysis 29

III. OUTCOME COMPARISONS AND PROCESS-OUTCOME RELATIONS: RESULTS WITH STATISTICAL ADJUSTMENTS 38
Effects of School Type on Student Outcomes 39
Relations between Processes and Achievement 48

IV. SUMMARY AND IMPLICATIONS OF THE EFFECTS OF ALTERNATIVE-SCHOOL PROGRAMS 67
Vouchers, Alternative-School Programs in Alum Rock, and Student Outcomes 67
Findings in Review 70
Implications of the Effects of Alternative-School Programs on Student Outcomes 78

Appendix

A. CORRELATIONS 83
B. MEANS, STANDARD DEVIATIONS, AND SAMPLE SIZES 84
C. READING SCORE DISTRIBUTIONS 86
D. PRELIMINARY ANALYSES OF THE FULL SET OF TEACHER SURVEY VARIABLES 92
E. TECHNICAL DETAILS OF THE ANALYSES 97

REFERENCES 103

TABLES

2.1. Number of Alum Rock Students with some Test Data for the period Fall 1972 to Fall 1976, by Year and School Type 15

2.2. Ethnic Composition of Alternative and Regular Schools During the Five Years of the Demonstration 16

2.3. Descriptive Statistics for Samples Used in Major Analyses ... 17

3.1. Analysis of MAT Reading Scores for Students in Alternative and Regular Schools 41

3.2. School Type-Reading Score Correlations at Pretest for One and Two Year Samples 43

3.3. Covariance Analysis of CSSCT Scores for Students in Alternative and Regular Schools 46

3.4. Correlational Results for Two Waves of Teacher Survey Variables (Teachers of Students with Data) 49

3.5. Regression of Class Level Student Reading Achievement on Teacher and Minischool Level Survey Variables 53

3.6. Effect of Program Size on Reading Achievement 60

3.7. Effects of Student Classroom Perceptions on Reading Achievement in the Fourth Year of the Demonstration: Standardized Regression Coefficients with Covariate Adjustment 61

3.8. Effects of Program Choice and Nonlocal School Attendance on Reading Achievement: Standardized Partial Regression Coefficients with Covariate Adjustment 65

A.1. Correlations of MAT Total Reading, Total Math, and Subtotal Math Scores 83

B.1. Total Sample Standard Score Means, Standard Deviations and Sample Sizes for Students in Alternative (A) and Regular (R) Schools 84

B.2. Analysis Sample Standard Score Means, Standard Deviations and Sample Sizes for Alternative (A) and Regular (R) Students . 85

D.1. Regressions of Spring 1975 Reading Scores on Individual and Aggregate Spring 1975 Teacher Survey Variables for Teachers and Students in Alternative Schools 93

TABLES (Cont.)

D.2. Regressions of Spring 1976 Reading Scores on Individual and Aggregate Spring 1976 Teacher Survey Variables for Teachers and Students in Alternative Schools 94

E.1. Index of Cross-References between Technical Outline and Text. 101

FIGURES

1.1. Schematic Representation of Information Examined in This Study 7

2.1 Mean MAT Reading Test (Standard) Scores, Fall 1974 to Fall 1976, for Cohorts of Students in Alternative and Regular Schools (Total Sample) 26

2.2 Mean MAT Reading Test (Standard) Scores, Fall 1974 to Fall 1976, for Cohorts of Students in Alternative and Regular Schools (Sample for Data Analysis) 28

C.1 Reading Scores--2nd Graders, Spring 1975, and 3rd Graders, Fall 1975 88

C.2 Reading Scores--3rd Graders, Spring 1975, and 4th Graders, Fall 1975 88

C.3 Reading Scores--4th Graders, Spring 1975, and 5th Graders, Fall 1975 89

C.4 Reading Scores--6th Graders, Spring 1975, and 7th Graders, Fall 1976 89

C.5 Reading Scores--2nd Graders, Spring 1975, and 3rd Graders, Fall 1975, Separately Within Alternative and Regular Schools 90

C.6 Reading Scores--3rd Graders, Spring 1975, and 4th Graders, Fall 1975, Separately Within Alternative and Regular Schools 93

I. INTRODUCTION

The Alum Rock Voucher Demonstration began in 1972 and informally ended in the spring of 1976.[1] During those four years the Alum Rock school district, with federal aid, experimented with a number of fiscal and administrative innovations in providing education to elementary students. Among these innovations were the decentralization of many areas of school decisionmaking; the creation of smaller, partially autonomous organizational units; and the establishment of mechanisms for increasing parent control over and participation in the schooling of their children.

The original intent of the demonstration was to implement a system of education vouchers, treating parents and students as consumers in an educational marketplace. Theoretically, in a voucher system the independent "producers" of schooling (the schools) receive vouchers equal in value to the district per pupil expenditure from each student choosing their "product." Schools responsive to consumer preference grow and expand, while unpopular schools can literally go out of business. All schools--public and private--compete for students.

At Alum Rock, however, several aspects of school finance and administration remained centralized at the district level, devaluing the voucher; formal constraints were placed on both the nature and degree of dynamic response to parent/student preferences; and private schools were not part of the demonstration. Moreover, not all schools in the Alum

[1]Technically, the demonstration was funded through 1976-77, but its most distinctive features were gone by June 1976. For details, see Bass (1978).

Rock Union Elementary School District participated in the demonstration. Participation was voluntary, and a little less than half of the schools in the district chose not to participate.[2] Within each participating school, two to five minischools were formed, each ostensibly teacher run and offering a distinct educational program.

While these features of the demonstration blur the relationship between the system of alternatives adopted in Alum Rock and the theoretical model of education vouchers, a number of innovations introduced into the district deserve investigation in their own right. The purpose of this report is to examine the effect of these innovations on students' reading achievement, social perceptions, and perceptions of themselves and their peers.

ROLE OF STUDENT OUTCOMES

Neither the theoretical model of education vouchers nor the modified system of alternatives implemented in Alum Rock directly attempted to improve the educational performance of the students served. Rather, benefits to students were expected to follow indirectly from the structural changes brought about by the demonstration, and from the reactions of parents and teachers to these changes. For example, students might reasonably be expected to benefit by a decentralization of decisionmaking, which would increase the teachers' freedom to explore new classroom practices and curricular materials. Or by having parents

[2]We call the schools participating in the demonstration "alternative" schools or "alternative programs." We call the schools not participating in the demonstration "regular schools." In other Rand reports, the former were usually termed "voucher schools" and the latter, "nonvoucher schools."

(and students) choose an educational program, students might be expected to benefit from the match of background to instruction. However, the success of the demonstration should not be judged primarily by student performance. If parents were more satisfied with their children's education, then the demonstration succeeded in meeting one of its stated objectives. The demonstration could not, of course, be judged completely successful if the performance of students in the alternative schools declined. A basic question investigated in this report is whether students profited as well as might have been expected if no innovations had been introduced.

In an examination of student outcomes, perhaps the most critical problem to solve is what to measure. As Snow (1974) points out, outcomes may be central, proximal, or distal. Central measures are linked directly to the content of instruction; they answer questions like "Did these students learn to identify the 50 states and their capitals on a map of the USA?" Such data, while valuable, are limited in two ways. First, since different programs have different objectives, often central outcomes cannot be used to compare different programs.[3] And second, given their limited scope, central measures do not comprise some of the important broader outcomes of an educational program. Proximal measures refer to broader measures of achievement in, say, reading and to measures of affect toward, say, self or others. Such measures are amenable to cross-program comparisons and probably

[3]Educational programs are not compared within an alternative school or between alternative schools. Rather, this report assumes that programs are homogeneous with respect to the cognitive and non-cognitive outcomes examined. Indeed, Barker, Bikson and Kimbrough (1981) found homogeneity of educational processes in classrooms.

correspond more closely to what parents and educators think of as outcomes than do the central measures. However, they are not as sensitive to program effects as are the central measures. Finally, the distal measures are broad indeed, covering such areas as learning to learn and general self-concept. For the purposes of this study, we used proximal measures of student outcomes. This permitted us to use common instruments to measure achievement and noncognitive outcomes in all classrooms of both traditional and alternative schools. However, a caveat is in order: proximal measures may be insensitive to some program effects. Hence, the conclusion should not be drawn that this report contains the final word on all outcomes of the demonstration.

FEATURES OF THE ALUM ROCK DEMONSTRATION AND THEIR RELATION TO STUDENT OUTCOMES

During the course of the Alum Rock demonstration a number of innovations were implemented to varying degrees in two stages: (1) moving from the "ideal" voucher demonstration to the Alum Rock model, and (2) modifying the Alum Rock model as it was put into practice. Differences between voucher theory and the Alum Rock demonstration are instructive with regard to the feasibility of the theoretical model of education vouchers. In the context of an analysis of student outcomes, these deviations from theory constitute insurmountable barriers to inferences about effects vouchers might have in practice. We may be justified in making statements about the effects of components of voucherized education which happen to have been implemented at Alum Rock, but such statements cannot speak to the issue of how these same

components would operate in a working voucher system. Thus, while some of the variables examined in our analyses might be integral parts of a voucher system, their effects cannot be taken as evidence for or against "vouchers."

In the second implementation stage, the program is seen as a list of specific innovations that vary in how closely they can be tied to voucher theory. Issues of the degree of implementation apply to the items on this list. For example, we know that smaller organizational units were created in the alternative schools; we are less sure, on the other hand, how extensively parents exercised their option to choose different alternative education programs for their children. Additionally, some of these innovations can be thought of as formalizing the provision of institutional support for practices already installed, but not widely used in Alum Rock (e.g., collaboration among teachers). Here we must distinguish between demonstration-related and naturally occurring variation in measured features of schooling in Alum Rock. Our aim is to view demonstration features--i.e., features of the alternative schools which differ from regular schools--as potential determinants of student outcomes. This goal is easily lost sight of when few variables are unequivocally treatment-related or setting-related, fully implemented or not implemented at all. But if our efforts are to be useful in guiding future thought on the merits of alternative innovations, it is desirable to make these distinctions where possible.

OVERVIEW OF THE RESEARCH

Our research attempts to provide answers to two broad questions regarding cognitive and noncognitive outcomes of the demonstration:

1. Did the demonstration affect student outcomes differently in alternative schools and regular schools?
2. Were particular features of the educational processes in these schools associated with variation in student outcomes? More specifically, "Were perceptions and attitudes of teachers, characteristics of the programs, perceptions of students, and choices of programs by parents associated with differences in student outcomes?"

A schematic of the types of information available for answering these questions is provided in Figure 1.1. Background variables included the relevant cognitive or noncognitive pretest (e.g., reading pretest for reading outcome), ethnicity, socioeconomic status (SES), language spoken in the student's home, and the number of days in school. Reading achievement served as the cognitive outcome; social, self, and peer perceptions served as the noncognitive outcomes. In the data analyses, the background variables served as covariates, and variation in the outcomes was adjusted for variation in background before program effects were estimated. These data permitted us to compare alternative schools as a whole with regular schools and thus to estimate the effect of the demonstration on cognitive and noncognitive outcomes.

In addition to being able to estimate overall program effects, we could use information on program processes in alternative schools to

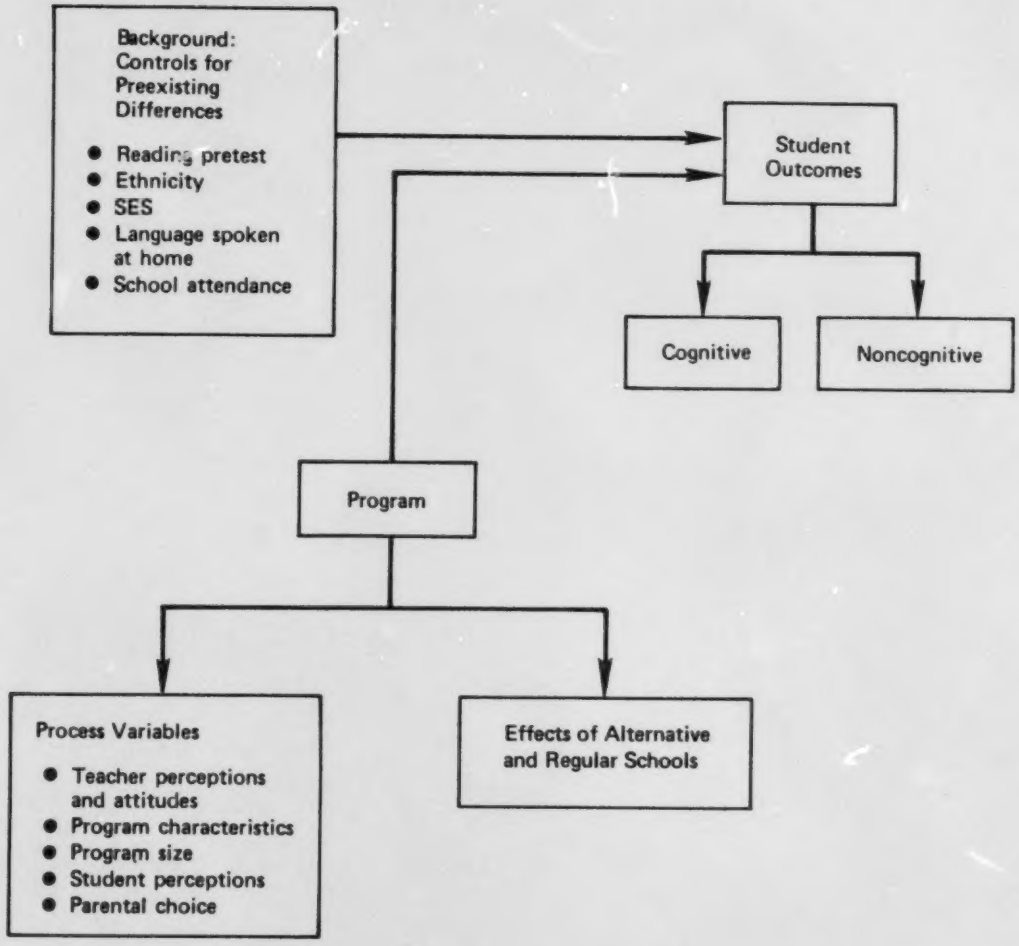


Fig. 1.1 — Schematic representation of information examined in this study

estimate the effects of these processes on students' reading achievement, after removing the effects of student background. One subset of process questions examined the relation between teachers' attitudes and perceptions and the performance of their students.[4] We can identify several dimensions of the school environment as seen by the teacher which might be influenced by the demonstration. The fiscal and administrative decentralization which occurred in alternative schools implies new roles for both teachers and principals. The smaller, teacher-run minischools should foster increased collaboration among teachers, perceptions of enhanced decisionmaking power, and a reduction in the extent to which the principal acts as a key figure in teachers' day-to-day affairs. Principals may be supportive of teachers' exercising increased responsibility, or hostile toward such a trend. Teachers' additional freedom of choice in curriculum areas may result in a sense that the work environment is conducive to experimentation with innovative approaches to teaching. The primary intent of our analyses is not to evaluate whether or not these changes occurred, but rather, if they occurred, to examine their effects on students. The teacher, then, acts as our observer on a number of aspects of what took place in Alum Rock.

A second type of information taken from the teacher surveys concerns their overall reactions to the demonstration: Did it have a positive effect on parents, on teachers, on students? What types of

---

[4]These data were taken from teacher surveys administered in the spring of the third and fourth years (1974-1975, 1975-1976) of the demonstration. For a detailed discussion of these teacher surveys, see Rasmussen (1981).

problems did teachers see as being created by the demonstration (e.g., student transfers, resource distribution, tension among minischools)? In relating these variables to student outcomes, the effects that reflect idiosyncratic characteristics of teachers must be distinguished from systematic effects that indicate demonstration impact. The former, which would probably have occurred whether or not the demonstration took place, are less relevant to an analysis of educational innovations. One possible approach to this problem would be to define the variables of interest at different levels of aggregation (e.g., teacher scores, minischool means on teacher scores) and to interpret effects at each level separately. For example, if all teachers in a minischool agree that there is tension between their program and another housed in the same school, the interpretation would be different than if a single teacher perceived such tension.

The creation of minischools was one of the most striking features of the demonstration; such teacher effects as are found occur in the demonstration-altered context of these smaller organizational units. But whether this structural change affected students in ways other than through its effect on teachers must be determined. That is, we examine whether minischool size is related to student achievement and whether this relationship is moderated by teacher perceptions and attitudes.

Roughly paralleling the inquiry into the effects of teacher survey variables on student outcomes is the question of the relationship between the students' own perceptions of the school environment and their performance on outcome measures. In the spring of the third and fourth years of the study, children in selected grades provided their

perceptions of the classrooms. Variables derived from these surveys are used to answer two questions: Do students in alternative and regular schools differ in their social perceptions of themselves and others? Are differences in classroom perceptions related to differences in achievement outcomes?

Another distinctive feature of the Alum Rock demonstration was that parents or students could choose different alternative-education programs in the demonstration without having to attend a local school. If such choices were made on the basis of information about the merits of various programs, or their suitability for particular children, the net result might be an improvement in the performance of children with this option over those without it. A more refined analysis would compare students/parents who exercise the choice option with their peers in alternative schools who do not.

A central difficulty in exploring the effects of choice arises when we attempt to determine whether a given choice was an educationally "informed" choice. If the decision to place a child in a particular minischool was made for reasons other than the nature of the education offered in that program (or in the program the child is leaving), we should hardly expect any educational benefits to accrue to that child. Conversely, we have no basis for distinguishing null choices (i.e., choosing to remain in the same program) which represent, say, apathy or lack of information from those representing an informed decision <u>not</u> to change programs. Thus, any effort to pinpoint the effects of student choice will be limited in scope and tentative in its conclusions.

Our approach to this question relies on two pieces of information about each child: 1) The child's history of schools and programs attended; and 2) the child's "catchment area" school, the school the child would attend if he/she were to attend the "local" school. With this information we can identify how frequently children changed schools or programs, how frequently they attended non-local schools, and cross-classifications of these two categories (e.g., attending a non-local school to remain in the same program when the family changed residence).

In summary, in addition to the quasi-experimental comparisons of the student outcomes (both cognitive and noncognitive) in alternative and regular schools, the following research questions are examined:

1. What were the effects of teachers' perceptions of and reactions to the demonstration on the performance of the students in their classes and in their programs and/or minischools?
2. How did the size of the new organizational units affect student performance? Were the effects of program size mediated by the teacher variables examined in the previous analyses?
3. Did differences in students' classroom perceptions reflect themselves in differences in performance on achievement outcome measures?
4. What impact did program choice options have on student outcomes? Were the effects different at different levels of aggregation?

SCOPE AND LIMITATIONS OF THE RESEARCH

The Alum Rock experiment in alternative education was formally operational from the fall of 1972 through the spring of 1976; during the fifth year of the demonstration many central features of the innovations were abandoned. The demonstration as a whole evolved considerably from its early days through its informal termination at the end of the fourth year. Rand followed the demonstration through its entire course, documenting the attitudes, reactions, and performance of the participants. As the demonstration progressed, survey instruments were adapted to be more sensitive to issues of emerging importance, and procedures for gathering data from children in the district were tailored to minimize objections from school personnel. In short, data collection decisions were influenced by many factors other than the desire to build a database supportive of sophisticated outcome-oriented analyses.

The analyses of student outcomes reported herein fall short of an exhaustive documentation of the progress of Alum Rock students, and the factors related to that progress. Virtually no achievement data were available from the first two years on students in regular schools, making comparisons impossible before the third year of the study. Those achievement data available on students in alternative schools from the first two years of the demonstration are flawed by test administration problems. Problems exist in the data from the later years of the study also. First, noncognitive measures were administered only to small and potentially unrepresentative subsamples of children. Second, only a

portion of the mathematics achievement test was administered, preventing the derivation of usable scores.[5]

OUTLINE OF THE REPORT

Chapter II describes the sample of students and an overview of the database, the process and the outcome measures used, and the methods employed in data analyses. We describe how the variables used in the analysis were constructed, and some of the technical characteristics of the scores. The sample descriptions are divided into two sections, one dealing with the "total" sample of students on our data tapes, and the second describing the subsample of students on which the major analyses are based.

Chapter III reports the findings of the study of student outcomes. In doing so, it also details the technical and practical problems that have determined the nature of the statistical analyses. The limitations of the analyses and a framework for the interpretation of the results are outlined.

Chapter IV presents a discussion of the findings, and attempts to formulate a set of conclusions about the effects of the demonstration on the students in Alum Rock. Implications for educational policy, questions raised by the findings, and areas for future investigations are outlined.

---

[5]The achievement test used in Alum Rock during the years of the demonstration covered in this report was the Metropolitan Achievement Test (MAT). The problems cited here largely resulted from difficulties in securing the agreement of Alum Rock teachers to a broad program of achievement testing specified by Rand.

II. METHODS

In this section we discuss sample characteristics, classifying students on the basis of family background and type of school attended; the derivation of scores; mean MAT reading achievement scores; and the statistical adjustments of scores.

DESCRIPTION OF THE SAMPLE

The Alum Rock Union Elementary School District is one of eleven school districts serving San Jose, California, in the sprawling metropolitan area south of San Francisco. The district serves a primarily residential area, lacking both industrial and major white collar employment centers. Much of the population is transient, with considerable mobility within the district. The general level of socioeconomic background of many Alum Rock residents is low. For example, in 1970, 10.4 percent of district families had incomes below the poverty level (compared to a statewide average of 8.4 percent).[1] The percentages of adults 25 years old or over with high school education ranged from 16 percent to 73 percent for the twelve census tracts in the district (two-thirds of the tracts containing fewer than 50 percent high school graduates). The population served by the district was about half Spanish surname (mostly Mexican-American), and about 12 percent black at the outset of the demonstration.

[1]U.S. Department of Commerce, 1970 Census of Population and Housing.

The district experienced sharp increases in total enrollment, as well as an increase in the relative size of the minority population, during the ten years preceding the demonstration. Over the five years of the study, however, total enrollment declined slightly (from about 15,400 to 13,800), but minority enrollment (in particular, Spanish-surnamed students) continued to increase. Schools in the district were, by and large, balanced with regard to ethnicity. Some imbalance was observed at the minischool level, apparently resulting from the multicultural and bilingual emphasis of particular programs.

Table 2.1 shows the numbers of students for whom data were available, classified by year of the demonstration and "school type" (i.e., regular or alternative). Virtually no data were available for students in regular schools prior to the third year of the study. Since our analysis focuses on the third and fourth years, we have omitted those students who were not in grades K through 8 during this period.

Table 2.1

NUMBER OF ALUM ROCK STUDENTS WITH SOME TEST DATA FOR THE PERIOD FALL 1972 TO FALL 1976, BY YEAR AND SCHOOL TYPE[a]

Year	Number of Students	
	Alternative School	Regular School
Fall 1972	3,167	268
Fall 1973	6,358	--
Fall 1974	6,010	3,379
Fall 1975	7,091	4,188
Fall 1976	5,733	3,612

[a]Entries in the table are counts drawn from the master data file for Alum Rock students, after omitting those not in grades K-8 at the beginning of the third year of the demonstration.

Table 2.2 divides the samples in Table 2.1 into percentages of students classified by ethnicity and school type for yearly cohorts from 1972-1976. During each year of the demonstration, traditional schools contained higher percentages of Spanish-surnamed students.

Table 2.3 provides descriptive information on the samples selected for the major analyses. We chose to focus on four grade-level cohorts with usable test data that were in the elementary grades throughout the third through the fifth years of the study.[2] The cohort sample sizes are approximate, since these vary from analysis to analysis. The samples are all roughly 60 percent Spanish surnamed and 60 percent free lunch eligible. Upwards of 60 percent of each sample are students in alternative schools.[3] The correlations in the table show the relation

Table 2.2

ETHNIC COMPOSITION OF ALTERNATIVE AND REGULAR SCHOOLS DURING THE FIVE YEARS OF THE DEMONSTRATION

	Percent Spanish		Percent Black		Percent Other	
Year	Alternative School	Regular School	Alternative School	Regular School	Alternative School	Regular School
Fall 1972	52	52	11	17	37	31
Fall 1973	56	--	17	--	27	--
Fall 1974	57	47	27	40	16	13
Fall 1975	59	47	25	39	16	14
Fall 1976	60	53	24	35	16	12

[2]Grades shown in the table are as of spring 1976.

[3]The grade 6 cohort contains a somewhat higher percentage since, by the sixth grade, many students in the traditional schools had begun to move to the middle schools which were predominantly demonstration schools.

Table 2.3

DESCRIPTIVE STATISTICS FOR SAMPLES USED IN MAJOR ANALYSES[a]

Grade[b]	N	Proportion Spanish Surnamed	Proportion Free Lunch Eligible	Proportion Alternative	Correlation: School-Type with Ethnicity[c]	Correlation: School-Type with Free Lunch[d]
3	683	.64	.57	.63	.020	.153
4	702	.60	.56	.61	.078	.211
5	780	.61	.59	.62	.113	.240
6	813	.59	.57	.75	.066	.197

[a]Sample drawn from same data file as those in previous tables.

[b]Grade level as of Spring 1976.

[c]Correlation (phi) between Spanish-surnamed (1 = yes; 0 = other ethnic group) dummy variable and school type (1 = alternative; 0 = regular). Indexes the disproportionality of the distribution of ethnic groups between alternative and regular schools.

[d]Correlation (phi) between Free Lunch Eligibility (1 = eligible; 0 = ineligible) and school type.

between school type and ethnicity, and between school type and eligibility for free lunch. The positive coefficients indicate that higher percentages of students who were Spanish surnamed or eligible for free lunch were in the demonstration schools. The samples selected for analysis contain slightly more Spanish-surname and free lunch eligible students than the total sample.

DEPENDENT VARIABLES

Cognitive (Achievement) Outcomes

The achievement data used in the analyses are total reading scores from the Metropolitan Achievement Test (MAT). Math scores were not included in our main analyses. Only two of the three math subtests were administered, and then only during 1974-1975, in order to reduce testing time in the schools. For comparative purposes, we present in Appendix A correlational analyses showing relations among math subtests, math total scores when they are present, estimated math total scores based on available subtests, and reading total scores.

Noncognitive Outcomes

The Children's Self-Social Construct Test (CSSCT) was administered to students to assess the demonstration's effect on a number of noncognitive, psychosocial constructs. These noncognitive outcomes are viewed as important both as outcomes in themselves and as potential facilitators of school achievement. The instrument, based on work by Ziller (1973), measured the dimensions of self-esteem, social distance

from significant others in the school environment, scope of peer attachment, social interest, perceived inclusion, and perceived individuation. Analyses of the CSSCT suggested that students in alternative schools "perceive themselves as closer to their teachers, more attached to their peers, and more generally included in the domain of social influence [than students in regular schools]; yet they retain a stronger sense of individuality" (Bikson, 1977).

Four measures were formed by averaging CSSCT items falling into subscales identified by Bikson:

1. Social Distance. The social distance subscale measured: (1) social distance from significant others in the school environment (i.e., from both teachers and peers) and (2) lack of social interest. For items measuring social distance, the student was shown a row of circles with one end marked with the target figure (teacher or peer), from which the student selected one to represent himself. "Social distance" was the distance the subject put between himself and the target figures. For items measuring lack of social interest, the student was shown a set of three circles (representing parents, teachers, friends) forming a social influence triangle, and asked to draw a circle representing himself/herself anywhere on the page. Lack of social interest items were scored as to whether the subject located himself/herself in or out of the triangle.
2. Self-Esteem. Self esteem was measured by having a child choose a circle to represent himself/herself in a vertical or

horizontal row of circles. Choice of a high circle from the column or a circle to the left end of the row indicated high self-esteem.

3. Peer Attachments. The subscale, peer attachments, was formed from items where a student could draw as many lines as he/she wanted, connecting circles representing other students to a circle representing himself/herself.
4. Individuation. Perceived individuation was measured by asking a student to choose a self circle from a collection of circles, a few of which were different from the majority of circles.

COVARIATES AND INDEPENDENT VARIABLES

Background Controls for Preexisting Difference

Background measures of ethnicity, language spoken at home, and socioeconomic status were used as covariates in the analyses. Since the major ethnic group consisted of students with Spanish surnames, a dichotomous variable indicating Spanish surname or not was constructed as a measure of ethnicity. For language spoken at home, a variable was created indicating whether Spanish was spoken at home. A student's eligibility for the federally funded free-lunch program was used as a proxy for SES, since determination of such eligibility is based on family income and size. We recognize potential sources of systematic error in the measurement of free-lunch eligibility. At the individual pupil level, all potentially eligible may not apply, thus inflating our estimate of SES for the sample, while at the school level, overreporting

of eligibility figures occurs (as these may form the basis for the distribution of other resources), biasing our estimate in the opposite direction. The free-lunch index was constructed from either third or fourth year data, depending on availability.

Absence from school was also used as a covariate in the analysis as a possible intermediate level program effect and moderator of achievement differences. The absence variable was defined as the sum of excused and unexcused absences during each demonstration year.

Process Variables: Teachers' Perceptions

To measure broad program characteristics such as organizational changes, teaching innovation, and diversity, Alum Rock teachers were surveyed periodically. A number of scales were formed from items on the survey instruments. These scales were empirically verified using factor analyses and were found to be similar to those from earlier analyses of the data.[4] Reliabilities (coefficient alpha) for all teacher survey scales used in this report ranged from .65 to .86.

The surveys of teacher opinion (spring 1975 and spring 1976) were designed to assess teachers' perceptions of their schools and attitudes toward the demonstration during its third and fourth year. Slightly different versions were designed for teachers in alternative and regular schools. The survey included items on specific topics such as attitudes toward the demonstration, perceptions of change in Alum Rock, school-community relations, minischool and school organization, diversity and its effects, decentralization and decisionmaking, economic incentives

[4] Rasmussen (1981).

and expenditure decisions, the school as a work environment, and teacher background.

Analyses of teacher survey responses yielded the following scales:

1. Overall effect of the demonstration--this scale was formed from questions referring to the effect of the demonstration on quality of education and on teachers.
2. Principal influence* [5] --This scale reflected the teachers' perceptions of the principal's influence on curriculum, budget, and new teacher hiring.
3. Teacher influence--This scale reflected teachers' perceptions of their influence on curriculum, budget, and new teacher hiring.

Teachers participating in the demonstration responded to demonstration-specific questions as well. Additional items were included in the "demonstration effect" scale (i.e., 1 above) and two additional scales were used:

4. Overall problems in the demonstration*--This scale was formed from questions about tension between minischools, enrollment instability, student and teacher transfer rules, and discretionary-fund allocation fairness.
5. Common policy--This scale examined the extent to which a common policy was shared by teachers at a minischool (regarding

[5] For asterisked (*) scales, high scores indicate "less"; for all other scales, higher scores imply "more" or "better."

learning objectives, teaching methods, behavior standards, grouping of students).

The Work Environment Scale (WES) was administered to teachers in 1975 and 1976 to measure perceived differences among the work environments of teachers in alternative and regular schools. Most of the 40 items on the WES were taken from a larger Work Environment Scale developed by Moos and Insel (1974). This scale has been used to describe the social climate of all types of work units, including basic organizational structure, direction of emphasis on personal growth and development, and interpersonal relations. A few items were developed by Rand staff to focus on frequently mentioned concerns of teachers in some alternative schools during the initial years of the demonstration.

WES items yielded scales indicative of the following aspects of the school work environment:

1. Staff cohesion --assessing the degree of group spirit and personal interest among staff members.
2. Principal support --measuring the extent to which the principal talks down to staff members or discourages criticism.
3. Teacher autonomy --describing whether teachers can use their own initiative and have freedom to do as they like.
4. Task orientation --measuring degree of organization, efficiency, work orientation.
5. Principal availability --indicating whether the principal is available and can be seen by staff members when the need arises.

6. Innovation --indicating the extent to which new and different ways of "doing things" are valued and/or tried.

7. Work pressure* --measuring the amount of pressure to keep working or the sense of task urgency.

8. Principal control* --indicating emphasis on policies and rules, and whether the principal keeps close watch on staff members.

Process Variables: Students' Perceptions

The Classroom Environment Survey (CES) sought students' descriptions of their classrooms with the goal of identifying learning environments that had a favorable effect on students. Survey items, drawn from other questionnaires, assessed students' perceptions of competitiveness, interpersonal friction, difficulty of their classwork, and adequacy of control of their classroom.

The following six scales were formed from the CES for use in the current analyses:

1. Liking --the extent to which the teacher is seen as having positive feelings about the students and whether the students like the class.
2. Organization --the degree to which the class is well-controlled and quiet with students busy (etc.).
3. Ease of class work --the level of ease of classroom tasks.
4. Competitiveness* --the degree to which students compete for high grades in the class.
5. Affiliation --the extent of positive social relationships among students.

6. Friction* --the amount of interpersonal conflict among students.

STUDENT PERFORMANCE ON THE READING TESTS

In this section we describe in general terms the performance of students in alternative and regular schools on the MAT reading achievement test during the last three years of the study. To simplify the presentation of data from different levels of the test, scores have been transformed to a common metric in the manner recommended by the test publishers. These "standard scores," which suffer from a number of technical shortcomings as discussed later in this chapter, are presented here to provide only a rough description of the relative performance of students in alternative and regular schools. Again, we caution the reader that these data have not been adjusted statistically to equate different groups of students. The analyses incorporating these statistical adjustments are reported in the next section.

Figure 2.1 shows mean reading achievement scores for all elementary grade children for whom test data were available. Each of nine separate cohorts of students in alternative and regular schools is represented by a white band. The width of the band indexes the mean difference in performance between the alternative and regular schools for the five testing sessions (fall 1974 to fall 1976). For example, the short band at the bottom right of the figure shows the performance of those students in alternative and regular schools who were in the second grade in the fall of the fifth year. The band in the upper left of the figure corresponds to children in the eighth grade in the third year.

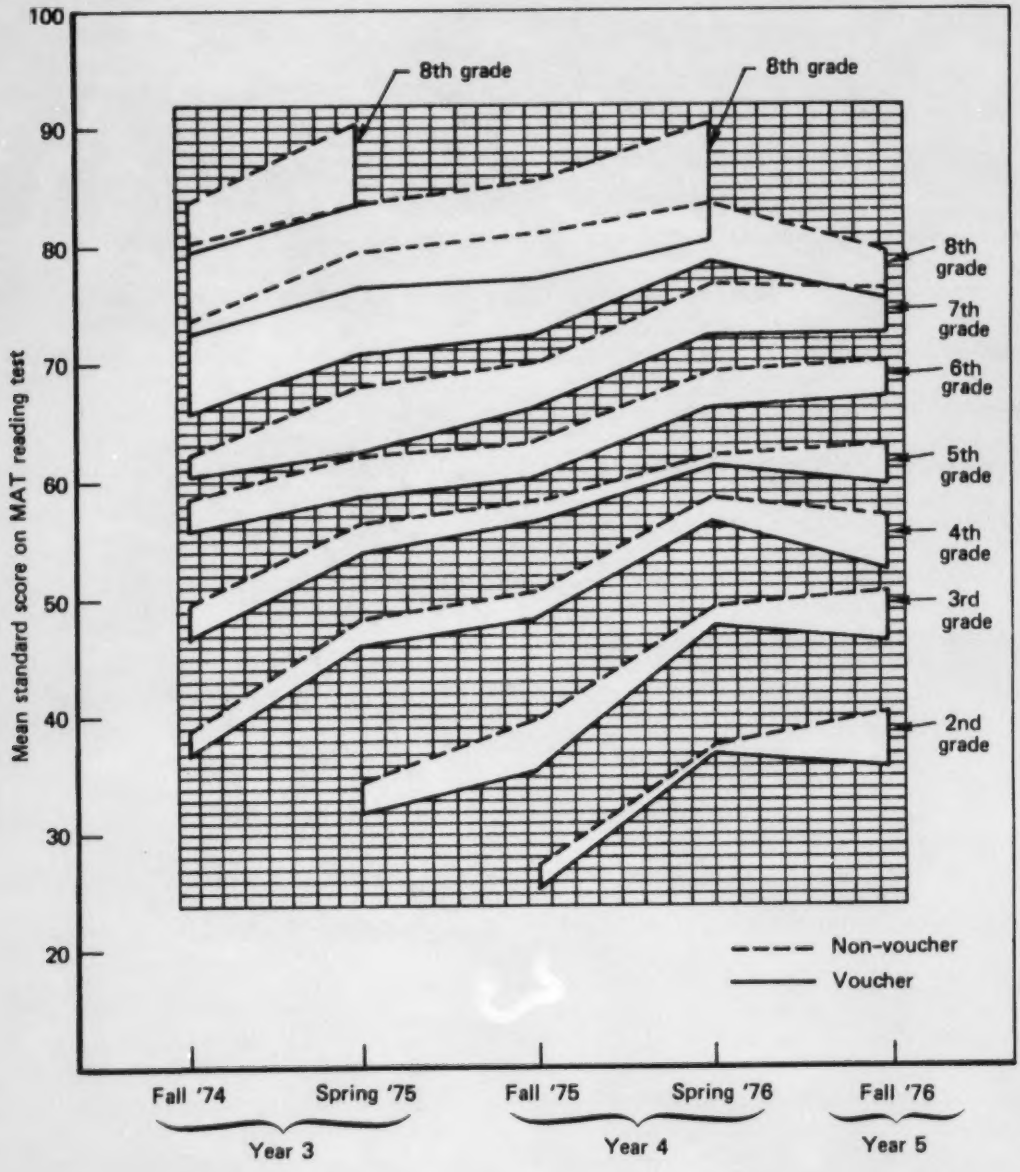


Fig. 2.1 — Mean MAT reading test (standard) scores, fall 1974 to fall 1976, for cohorts of students in alternative and regular schools (total sample)

The students in regular schools consistently outperformed students in alternative schools--the regular students' scores are the upper edge of each band--but the figure shows no trend for the gap between school types to increase over time. This is especially noteworthy since the variability within groups increases over time (e.g., the standard deviations for sixth graders in alternative schools are larger than the corresponding statistics for fifth graders in alternative schools), and in the absence of intervention, one would expect differences in the means of the two types of schools to exhibit this same trend. There does appear to be a slight drop in the performance of students in alternative schools in the lower grade levels from the end of the fourth year to the beginning of the fifth year. This decline probably was not due to the demonstration, since the demonstration informally ended in the fourth year. In addition, the cohort for which the decline is most pronounced, fourth graders in fall 1976, received a more difficult level of the MAT than was given to fourth graders in previous years. The inadequacy of the transformation to the standard score metric may be operating to produce a spurious difference in the group means for this cohort.

Figure 2.2 presents data similar to those presented in Figure 2.1 for the subsamples of students on whom the statistical analyses were conducted. (The tables in Appendix B report means, standard deviations, and sample sizes corresponding to Figures 2.1 and 2.2.)

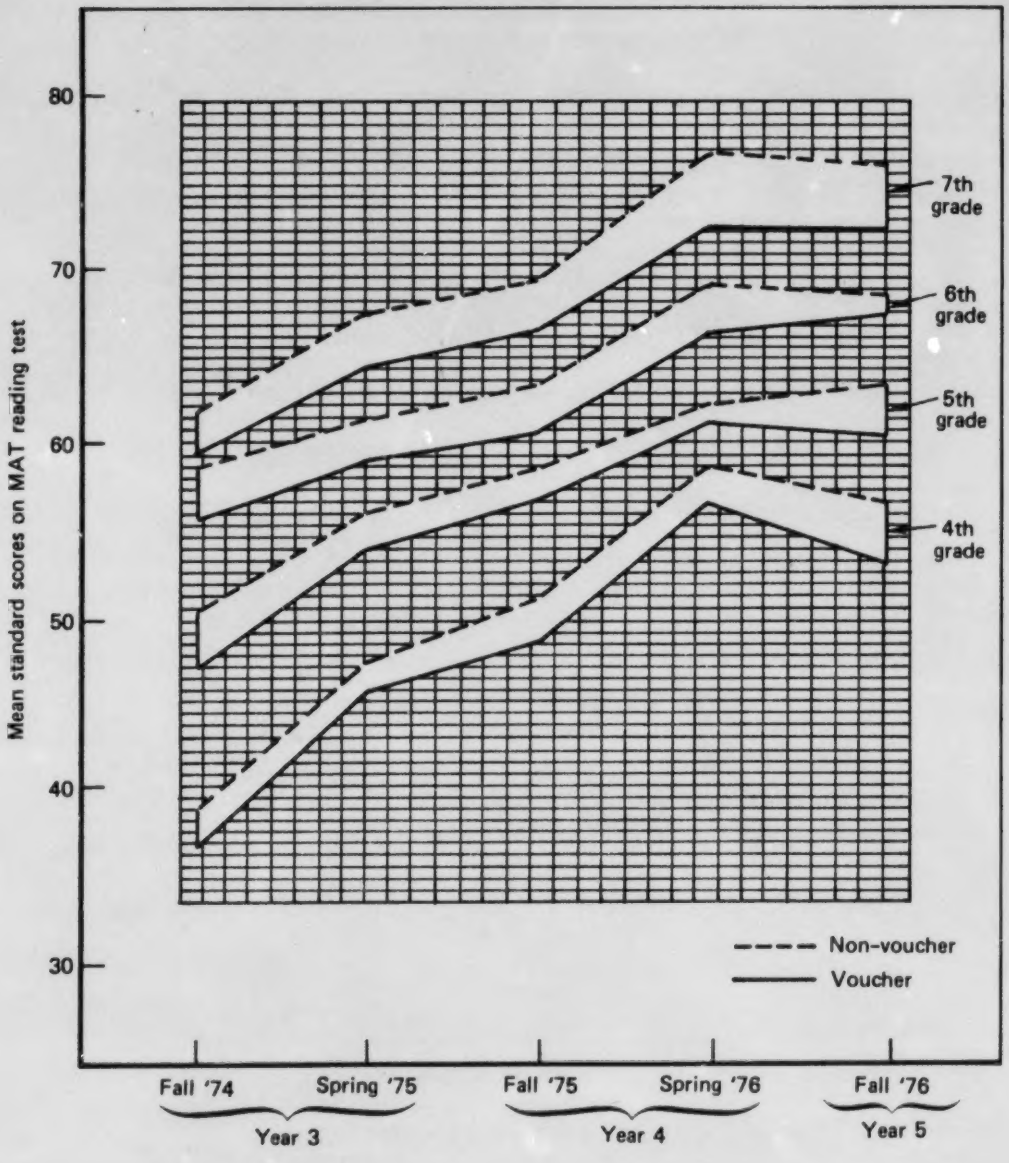


Fig. 2.2 — Mean MAT reading test (standard) scores, fall 1974 to fall 1976, for cohorts of students in alternative and regular schools (sample for data analysis)

DATA ANALYSIS

Comparisons of Alternative and Regular Schools

Since schools self-selected into the demonstration, students and teachers were not randomly assigned to levels of the school type variable. The design of the study, then, may be characterized as a nonequivalent control group design, i.e., a quasiexperiment. This means that differences between the two groups may be due to the effects of the demonstration, to prior existing differences in the two groups, or to some combination of the two. Statistical adjustments must be made to rule out prior existing differences as an explanation for the observed differences between students in alternative and regular schools. These statistical adjustments attempt to create a situation where "all other things are equal."

No standard procedure is available for determining the appropriate adjustment in any given application of the nonequivalent control group design (Lord, 1967; Cronbach, Rogosa, Floden and Price, 1977). The difficulties are both practical and logical/theoretical. Groups may differ in many ways, and it is seldom possible to obtain all the necessary data for a "complete" adjustment. Also, preexisting group differences may interact with treatment characteristics in a more complex manner than is assumed by adjustment procedures (e.g., nonlinear or discontinuous relations). In summary, nonequivalent control group comparisons rely on incomplete or proxy information about preexisting differences, and on simplifying assumptions concerning how such differences may affect outcomes. Given these limitations, the

interpretation of adjusted group differences as reflections of treatment effects is tenuous at best.

One way to increase the likelihood of obtaining reasonable answers to questions of program impact with nonequivalent groups is to examine the data from several analytic perspectives, each perspective making different assumptions (Wortman and St. Pierre, 1977). As diverse analyses converge on similar findings, we may venture interpretations and conclusions with greater confidence. This approach is, admittedly, only a crude approximation to complete delineation of the processes that are causing outcomes, since each method embodies a substantively different model.

Several features of the analyses strengthen the nature of interpretations which may be given to the results. First, for reasons detailed below, the analyses are run on four (or three, depending on the analysis) separate cohorts of children, defined with respect to grade level. If we find consistency across cohorts in outcomes, we can be more confident that the effects are not restricted to a single, perhaps atypical, group of students. Second, the Alum Rock dataset is longitudinal, containing multiple waves of usable data. This feature makes it possible to assess trends in outcomes and the stability of effects over a longer period than in many studies. Third, multiple analysis models are employed--analysis of <u>covariance</u>, using one (pretest) or a full set of background covariates, with and without correction for unreliability in the pretest. While these methods in no sense exhaust the range of possible techniques, they do represent distinctly different models, and robustness of findings across the four

would provide a relatively sound basis for interpretation of demonstration effects.

Analyses of Process Variables

The second phase in the analyses involves a closer look at how changes adopted in participating schools affect student performance. Did variations in the activities and climates in alternative schools affect students' achievement? If so, how?

Schools can affect students in different ways. The principal controls aspects of the school that affect all who work or study there. A classroom teacher primarily affects only those students in her class, but she also makes a contribution to the work climate in the school. Process variables, then, can be associated with different levels of aggregation--those that vary from one student to the next, from one classroom to the next, one minischool to the next, and so on. Similarly, the innovations introduced at Alum Rock stand out at different levels of aggregation, and the analyses of their effects need to take this into account.

Participants in the demonstration (especially teachers) can respond to structural changes or to increased professional freedoms and incentives in a variety of ways, ranging from individual rejection of new trends to unanimous acceptance. Of course, some reactions are not based on careful consideration of the program's merits; instead they reflect indiscriminate predispositions (e.g., unbridled enthusiasm, chronic pessimism). In addition, participants influence one another, and the opinions and practices of groups become formal policies whose

enforcement may vary from one individual (e.g., a minischool staff member) to the next.

While the literal content of teacher survey questions dealt with the distribution of authority and the perceived organizational climate, it is possible that individual and aggregate response summaries can differ in meaning. An aggregate measure of perceived influence, for example, should index the level of influence common to all teachers in the minischool; minischools with high aggregate scores presumably have transferred more authority to teachers than those with low scores. In contrast, how a given teacher's view differs from the minischool consensus probably conveys less information about the extent of decentralization characterizing the minischool. Rather, differences among teachers within minischools may well reflect attitudinal or other predispostions likely to persist even in the absence of the decentralization program. In other words, an individual teacher's response to the survey might plausibly be influenced by two distinct sources of variation: (1) the degree of implementation of the program component (in this case, decentralization of decisionmaking) and (2) individual differences among teachers in their personal characteristics and/or their orientation toward their work.

It is unclear, then, whether an analysis using survey data at the individual teacher level would tell us about how the redistribution of decisionmaking authority and other features of the program at Alum Rock affected the students. But by decomposing individual teacher survey variables into independent components corresponding to the between and within minischool influences described above, it would be possible to

separately examine their effects on students. The two components of teachers' survey responses need not exert consistent influences on student outcomes. If, for example, decentralization meant increased time on noninstructional activities for teachers, or the novelty of the program were otherwise disruptive, the aggregate effect on students might be negative. Alternatively, if a highly efficacious teacher reports greater perceived influence than his or her peers in the same minischool--whether or not the minischool has implemented the program--the relative effect within schools of the teacher influence variable might well be positive.

Other aspects of the Alum Rock program, such as parent options to choose among alternative educational programs, are approached in a way similar to the approach for the process variables from the teacher survey. Again, the effects at different levels of aggregation may take on somewhat different meanings. For example, while program changes on the part of an individual student (i.e., movement from one program to another) may be beneficial for that student, classrooms or programs with very high student turnover rates may constitute less than optimal learning environments. Thus, it is important to examine such variables at different levels of aggregation before settling on an interpretation of their effects.

The purpose of the analysis is to provide separate estimates of the effects of the process variables at multiple levels.[6] To see how this

[6] This analysis model provides conservative estimates of the adjusted group level effects. Controversy exists over a number of alternative formulations of the multilevel analysis problem. See Burstein (1980) and Firebaugh (1979) for additional discussion of the technical issues.

is accomplished, consider the following simple equation, where student outcome, y, is regressed on a single teacher variable, T, and its minischool level aggregate $\bar{T}$:

$$y = \beta_1 T + \beta_2 \bar{T} + u$$

This is a simplified version of the actual analysis and the basic form of models used to estimate "contextual" effects (Alwin, 1976; Alwin and Otto, 1977; Burstein, 1980; Hauser, 1971; Lincoln and Zeitz, 1979). The interpretation given to the two regression coefficients β_1 and β_2 is as follows: β_1 is the effect of the teacher level variable, holding constant the effect of its minischool level aggregate, i.e., it reflects the magnitude of change in y associated with a one unit change in the classroom teacher's relative standing within his/her minischool. Similarly, β_2 is the effect for the minischool level aggregate of the teacher variable, holding constant the individual teacher level variable, T; it is the contextual or "structural" effect. Note that β_1 and β_2 are related to the total effect for the variable T (i.e., the coefficient from the regression excluding $\bar{T}$ from the equation) by the following expression:

$$\beta_{yT} = \beta_1 + \eta_T^2 \beta_2,$$

where η_T^2 is the ratio of between-minischool to total variation in T, and β_{yT} is the coefficient from the usual y on T regression (Duncan, Cuzzort, and Duncan, 1961). Clearly, if β_1 and β_2 index different substantive processes, β_{yT} will be an uninformative mixture of the two distinct effects (Cronbach, 1976).

Complexities in the Data Structure

The Alum Rock dataset is longitudinal, and membership in groups (classrooms, programs, etc.) changes over time, resulting in an ambiguity in the specification of a grouping rule for any given variable (Dyer, Linn, and Patton, 1969). In the analyses reported below, aggregates are constructed according to group membership at the time of measurement of the outcome variable. This procedure mixes between- and within-group variation in data collected prior to the formation of the "current" groups, making the assumption that individuals carry the effects of previous contexts with them into the new groupings. There seems to be no completely satisfactory approach to this problem when, for example, the variable of interest refers to movement from one aggregate to another.

We decided to conduct some of the student outcome analyses separately by grade levels because of the undependability of the test publisher's vertical equating for Alum Rock students. This has implications for how the data may be interpreted. The effects of program-level (and to some extent class-level) variables should be expected to overlap from cohort to cohort since students in different grades (and consequently in different analysis samples) within a given program receive the same value of any program-level variable. Also, where classrooms contain students in multiple grade levels, the within grade level analysis will place students from the same classroom into different analysis samples. Differential effects might be associated with relative grade level standing within a class or program (e.g., being a third grader among third, fourth and fifth graders, etc.). A

detailed examination of this type of hypothesis, however, is beyond the scope of the present effort. The general point here is that the separate analyses of grade level cohorts do not in any sense constitute independent replications--constraints on the form of the analyses may operate to create (possibly artifactual) consistency and/or inconsistency in the results across cohorts.

Missing Data

The dataset is incomplete in a number of ways. Non-response in the sample, the need to limit data collection, shifting emphasis in research goals as the demonstration unfolded, and the technical difficulties of coordinating a complex, large-scale database may all lead to missing data. The dataset has several major shortcomings:

- For several of the instruments data gathering was deliberately limited to subsamples of the Alum Rock "population." Included among these are the noncognitive measures administered to the students, and both teacher self-reports and independent observations of classroom practices.
- A rather general problem is encountered in any attempt to link data from a number of different sources (e.g., student test scores, family residence information, teacher survey responses). Links are established through correspondences in the identifying codes from the various sources. A missing code, such as a teacher ID or school code, results in an incomplete protocol for a student. Often the information used to match pieces of the data file was not ideally suited for

this purpose (e.g., to identify a student's teacher at a given point in time, we had to rely on the teacher ID coded on his/her achievement test form, which only indicates the teacher who administered the test, not necessarily the child's classroom teacher). Time and resource constraints have limited our systematic exploration of these problems, with the probable result of a less complete database than might be obtained with more extensive "detective" work.

- Complete and partial nonresponse on the part of subjects also reduces the size of the database. Nonresponse can be a function of the respondent, the information sought, or the measuring procedure itself. Teachers who return partially completed surveys may differ in important ways from those who do not. Questionnaire items probing sensitive issues may elicit lower response rates, as may poorly worded items. Again, we have achieved no definitive resolution of these problems insofar as they affect the database.

III. OUTCOME COMPARISONS AND PROCESS-OUTCOME RELATIONS: RESULTS WITH STATISTICAL ADJUSTMENTS

Has the demonstration modified the performance of the students in alternative schools as compared to students in regular schools? What features of the demonstration can be shown to account for variations in student outcomes?

In approaching the first question, the strength of the association between school type (i.e., alternative or regular school) and outcome scores is estimated after the questions have been adjusted statistically to remove group differences due to nontreatment related factors.[1] As discussed in Chapter II, the adjust procedure is not guaranteed to adjust correctly for systematic pretreatment nonequivalence of groups. Underadjustment is likely, but overadjustment is also possible. To guard against an incorrect conclusion about school type differences, the effects of school type are estimated using several distinct models. (For technical details of these models, see Appendix E.)

We approach the second question by exploring relations between features of the demonstration (e.g., scales derived from teacher surveys, student attendance changes, classroom climate) and the performance of participating students. Our analyses provide statistical

[1] Control variables used in forming the adjustments include ethnicity (Spanish-surnamed vs. other), Spanish language spoken in the home, eligibility for free lunch, number of days absent during the period spanned by the analysis, and the relevant premeasure. These variables do not exhaust the list of those on which the groups might differ. We proceed under the assumption that the variables we have ignored would not significantly distort our findings.

adjustments which are applied so that the aspects of schooling which serve as independent variables are not given explanatory power due to their relation to pre-treatment student characteristics. (For details, see Appendix E.) This is a conservative approach to the assessment of schooling effects since, to some extent, the covariation between class and/or program measures and student background is itself a demonstration effect. The general issue of how to partition background and schooling effects is more a matter of theoretical orientation than of statistical sophistication; effects found under our procedure run little risk of spurious dependence on nonschool factors.

EFFECTS OF SCHOOL TYPE ON STUDENT OUTCOMES

MAT Reading Outcomes

Four cohorts of students--3rd, 4th, 5th, and 6th graders as of spring 1976--and 3 time spans--fall 1974 to spring 1975, fall 1974 to spring 1976, and fall 1974 to fall 1976--are represented in the analyses. The fall 1974 administration of the MAT Reading test serves as the pretest.

The four adjustment procedures employed in the analyses vary in the extent of their statistical control. The simplest model, pretest-only, includes only the MAT Reading pretest in the covariate set. The second adjustment, multiple-covariate, includes the full set of covariates (see Covariates and Independent Variables in Chapter II; see also Footnote 1 in Chapter III). The third and fourth adjustment models are the same as the first two with the exception that the MAT Reading pretest is

corrected for measurement error.[2] The pretest-only model and the corrected-multiple-covariate model represent extremes in adjustment and are reasonably interpreted as upper and lower bounds, respectively, on the adjusted schooltype effects.

Table 3.1 presents the results for these four analyses. The first column in the table contains the correlations between treatment group membership (coded 1=alternative school at posttest; 0=not) and reading posttest scores. These coefficients index the unadjusted mean outcome differences between the two groups of students. The correlations are all negative (this information is roughly analogous to that in Figure 2.2), indicating that participating students perform more poorly. All but one of the schooltype-effect conditions, that for the middle time span 4th grade analysis, achieve statistical significance ($\alpha = .05$).

The next four columns provide estimates of the adjusted schooltype effects for the four adjustment procedures. The adjustments are formed on the basis of the pooled within-groups regression of outcomes on covariates.[3] The entries in the table are standardized partial regression coefficients for the schooltype variable. Positive values indicate an effect favoring alternative schools; negative values, the reverse.

---

[2] Reliability estimates were obtained separately for students in participating and nonparticipating school samples. Since item data were not available, we have used as a rough indicator of reliability the correlation between successive spring to fall administrations of a given test level to the same sample of students. These estimates ranged from nonparticipating students. Test score distributions can be found in Appendix C.

[3] The ANCOVA assumption of homogeneity of within-group regressions was examined by constructing the covariance adjustments separately for students in alternative and regular cohorts. Only modest deviations from this condition were found.

Table 3.1

ANALYSIS OF MAT READING SCORES FOR STUDENTS IN ALTERNATIVE AND REGULAR SCHOOLS

Grade/Year	Schooltype-Reading Correlation	Standardized Partial Regression Coefficients: Effect of Schooltype on Reading				N	Percent Alternative
		Pretest	Covariate Set	Pretest Adj.[a]	Covariate Set Adj.[a]		
Grade 3							
F74-S75	-.086*[b]	-.010	.003	.011	.023	626	61.2
F74-S76	-.084*	.000	.028	.022	.049	683	63.1
F74-F76	-.119*	-.021	-.010	.004	.015	577	63.1
Grade 4							
F74-S75	-.137*	.000	.020	.036	.055*	653	60.2
F74-S76	-.051	.046	.083*	.072*	.107*	702	60.8
F74-F76	-.138*	-.013	.000	.021	.037	636	60.1
Grade 5							
F74-S75	-.141*	-.001	.013	.023	.037	728	60.9
F74-S76	-.180*	-.040	-.024	-.016	.000	780	62.2
F74-F76	-.123*	.008	.024	.031	.047	686	62.7
Grade 6							
F74-S75	-.137*	-.056*[c]	-.050*	-.048*	-.043*	740	60.8
F74-S76	-.190*	-.073*	-.044*	-.062*	-.033	813	75.4
F74-F76	-.136*	-.018	-.007	-.006	.019	706	74.9

[a] Indicates correction for pretest unreliability has been applied.

[b] $\alpha = .05$.

[c] Coefficients are asterisked if the corresponding unstandardized coefficients exceed twice their standard errors.

First, note the overall reduction in magnitude of the adjusted effects in comparison to the posttest schooltype-effect correlations. Second, those effects remaining after adjustment do not consistently favor either alternative or regular schools. Third, the different adjustment procedures are ranked consistently in the magnitude of effect estimates, as expected. Not surprisingly, the corrected multiple covariate analysis produces the largest adjustments and the uncorrected pretest-only analysis adjustments are the least severe. In general, after correcting for nonequivalence of the two groups, treatment effects may be conservatively summarized as random fluctuation about zero. We find a slightly greater number of significant coefficients than would be expected due to chance, but our attribution of significance can be criticized as extremely liberal. When treatments are delivered to intact groups (e.g., classrooms and schools), individuals within these groups are not statistically independent units (Cronbach, 1976). Using the number of students as a basis for degrees of freedom in significance testing overestimates (in direct proportion to group size) the apparent importance of minor deviations from the null case.

A comparison of the coefficients for the treatment effect after one and two years for each cohort shows, in most cases, a shift toward a more positive effect of the demonstration (Table 3.1). This is the pattern we would expect if increasing amounts of exposure to a "treatment" were beneficial for students. If so, we might expect the trend to be absent in the longest time span both because the fall 1976 posttesting is likely to reflect decline over the summer, and because the dismantling of the demonstration had begun by this time.

An alternative account of this trend can be advanced. Inspection of the pretest "treatment-effect" correlations (see Table 3.2) reveals variability from one sample to the next. In particular, the magnitude of the coefficients for the samples included in the two year analyses are routinely greater in a negative direction than the corresponding correlations computed on the one year analysis samples. This indicates that when two-year longitudinal data are analyzed, there is a relatively wider gap in entering ability between the students in alternative and regular schools. A check on the pretest means (see Appendix Table B.2) credits the increased separation to downward changes in the composition of the alternative-school samples.[4]

Table 3.2

SCHOOL TYPE-READING SCORE CORRELATIONS AT PRETEST FOR ONE AND TWO YEAR SAMPLES

	Correlations	
Grade	One Year Sample (F74-S75)	Two Year Sample (F74-S76)
Grade 3	-.098	-.159
Grade 4	-.161	-.128
Grade 5	-.166	-.178
Grade 6	-.093	-.143

[4] For 6th graders the change is masked by the movement of students in regular schools to the alternative schools which dominate the middle schools. The competing explanation acknowledges a tendency (regression artifact) for students with comparatively extreme initial scores constituting the two year analysis sample to become less divergent at a later measurement occasion, even in the absence of any

In summary, the quasiexperimental comparisons of the MAT reading scores of participating and nonparticipating students indicate the absence of any positive or negative effect associated with varying amounts of exposure to the demonstration.

Noncognitive Outcomes

When we examine how students' self- and social-perceptions were influenced by the demonstration, we find several noteworthy differences between the analyses of noncognitive outcomes and those for the achievement measures. First, while the basic pre- to posttest design is followed, the data were gathered in the spring of the third and fourth years of the demonstration. Our examination of treatment effects, then, is limited to this interval. Second, grade levels are combined in the analysis since the CSSCT is appropriate for multiple grades. Third, in addition to examining the overall effect of school type on noncognitive outcomes, we test the possibility of a school type by grade level interaction, perhaps a more realistic model of where effects might be found.

There is also a more fundamental difference in the interpretation of the results of our analyses. In contrast to cognitive constructs, the interpretation of changes in the noncognitive domain as "growth" or "progress" is problematic. General consensus as to positive global directions for change (e.g., increasing "sociability") can be reached, but alternative patterns, deemphasizing or compensating for specific

---

intervention. The ANCOVA procedure is believed to be especially prone to producing biased adjustments in such situations.

social weaknesses, can always be constructed. In the same vein, it is often a matter of the data analyst's preference whether a positive or negative valence is attached to particular variables. Do we speak of "dependency" or "connectedness," "isolation" or "independence," "withdrawal" or "introspection?" To base such distinctions on data, refined measurement is required (e.g., multitrait-multimethod validation). As a pragmatic resolution of this problem, we employ the interpretive framework of CSSCT, recognizing the possibility that the same data could be used to support other construct interpretations.

Table 3.3 reports the results of the covariance analyses of the scores derived from the CSSCT.[5] In this and subsequent analyses our general approach is to use one adjustment procedure, namely a "multiple covariate" analysis without reliability correction. This approach, based on the outcome of the achievement analyses, seems least likely to produce over- or underadjustment.

The schooltype-effect correlations show students in alternative and regular schools, on average, becoming more similar from the third to the fourth years of the study (columns 1 and 2 of Table 3.3). There appears to be an increase in social distance and a reduction of self-esteem and scope of peer attachments for students in alternative relative to regular schools. The signs of the pretest treatment-effect correlations (i.e., negative for SOCIAL, and positive for ESTEEM and PEER) are consistent with the findings of Bikson (1977; see our discussion in Chapter II).

---

[5] We do not present analyses for the fourth scale derived from the CSSCT, "INDIVIDUATION," due to a problem in the distribution of this variable for our sample (zero variance) at the pretest administration.

Table 3.3

COVARIANCE ANALYSIS OF CSSCT SCORES FOR STUDENTS IN ALTERNATIVE AND REGULAR SCHOOLS

Variable	Correlations: Pretest with Schooltype	Correlations: Posttest with Schooltype	Standardized Partial Regression Coefficients: Schooltype[a]	Schooltype x Grade[b]	Grade[c]	N	Percent Demonstration
SOCIAL	-.085	.036	.066*[d]	.080*	-.103*	1130	65
ESTEEM	.056	.019	-.019	-.010	.008	1141	65
PEER	.086	.037	.021	.033	-.000	1119	66

[a]Standardized partial regression coefficient for "Schooltype" (1=alternative; 2= regular) holding constant the background covariate set.

[b]Standardized partial regression coefficient for the Schooltype by grade level interaction variable holding constant grade level and the background set.

[c]Standardized partial regression coefficient for grade level holding constant Schooltype by grade level and the background set.

[d]Coefficients are asterisked if the corresponding unstandardized coefficients exceed twice their standard errors.

The effect of school type, schooltype-grade interaction, and grade level on the noncognitive outcomes are reported in the next three columns of Table 3.3. More specifically, column three contains the estimated schooltype effect adjusted for covariates and other independent variables; column four contains the estimated effect of the treatment by grade-level interaction on noncognitive outcomes, adjusted for all other variables; and column five presents the effect of grade-level on noncognitive outcomes, adjusted for all other variables. For the variables ESTEEM and PEER, the effects of school type, grade level, and the school type by grade level interaction are negligible after statistical adjustment. With respect to SOCIAL DISTANCE, students in alternative schools are more distant from significant others than students in regular schools, more so than would be predicted from initial status. As a trend over time, this result takes on increased significance in relation to the general grade level effect in the opposite direction. While the district-wide tendency is for students in later grades to become socially less distant,[6] participating students (particularly those in later grades) have, during the period of our analysis, come to perceive greater distance between themselves and others. Without a clear account of the factors combining to produce variation in premeasures, the third to fourth year changes might be due to a negative demonstration impact, a decline after initial positive impact (after the first three years), or inadequate statistical methods.

[6] Here we make the plausible inference from a cross-sectional trend (i.e., the grade effect) to a longitudinal argument.

We should point out in concluding that the magnitude of the effects discussed here is small.

RELATIONS BETWEEN PROCESSES AND ACHIEVEMENT

Although the analyses in the previous sections, particularly the MAT reading achievement results, show no dramatic effects associated with attendance at alternative schools, we may still expect to find specific demonstration-related aspects of schooling to have had some impact on Alum Rock students. The intent in exploring relationships in this "zero sum" context is to fill in some of the detail lost in overall schooltype comparisons. Whereas in the ANCOVAs a categorical variable indexing schooltype is entered into a regression equation and evaluated, the analyses in the remainder of this section replace this static indicator with variables or sets of variables corresponding to features of the alternative schools. These variables come closer to capturing demonstration-related processes.

Teacher Survey Variables and Their Effect on Reading Scores

We view the assessment of teacher effects as a two stage process. Teachers change their behavior (or report changes in their environment) as a result of the demonstration. This change then influences their pupils. We present analyses bearing on these two stages in turn.

Table 3.4 contains indices of the demonstration's impact on teacher-survey responses during the third and fourth years of the program. The schooltype-effect correlations suggest that many teacher variables may be sensitive indicators of demonstration-induced changes in education. Only a single scale, principal control, fails to

Table 3.4

CORRELATIONAL RESULTS FOR TWO WAVES OF TEACHER SURVEY VARIABLES (TEACHERS OF STUDENTS WITH DATA)

Variables	Schooltype-Reading Correlations[a]		Stability Correlations[b]	
	1975	1976	All Teachers	Participating Teachers
Cohesion[c]	.144**	.288**	.519	.383
Principal Support	-.100*	.088	.326	.446
Autonomy	-.054	.121*	.313	.401
Task Oriented	.106*	.086	.515	.594
Principal Availability	-.253***	.070	.509	.584
Pressure	-.178***	-.042	.608	.542
Principal Control	.019	.040	.553	.515
Innovation	.047	.120*	.226	.299
Demonstration Effect	.389***	.477***	.686	.659
Principal Influence	.427***	.061	.326	.294
Teacher Influence	.471***	.305***	.502	.454
Demo Problems[d]	---	---	---	.589
Common Policy[d]	---	---	---	.540

[a]Correlations of scales with Group (1 = participating; 0 = not). Sample sizes range from $n = 247$ to $n = 351$. *:$p \leq .05$; **:$p \leq .01$; ***$p \leq .001$.

[b]Correlations between measurement occasions of the same variable. Sample sizes range from $n = 217$ to $n = 265$ for total sample correlations, and from $n = 125$ to $n = 152$ for participating correlations. All coefficients are significant.

[c]See variable descriptions in Chapter II.

[d]Defined for participating teachers only.

differentiate between teachers in alternative and regular schools at either measurement occasion.

To attempt a characterization of the teachers on the basis of these findings would be a lengthy digression; we mention here only what seem to be important changes from the third to the fourth years of the demonstration. During this period, teachers in alternative schools perceived increases in cohesion, autonomy, innovativeness, pressure in the work environment, and in the influence, availability, and support of their principals. At the same time, they perceived a reduction in their own influence. Overall, their sense of the worth of the demonstration was strengthened between the third and fourth years. Unequivocal interpretation of such changes is not possible without knowledge of the processes influencing the baseline values (e.g., drawing a sample of extreme initial scores vs. shifting patterns of demonstration impact on teachers).

Table 3.4 also shows the stability of teachers' reports on the features of the school environment. Presented side by side are one-year stability correlations corresponding to the total sample and alternative-school teacher sample. If, during this interval, intervening events in alternative schools were responsible for some major upheaval in the perceptions of teachers, we should expect to see reduced stability (i.e., lower correlations) for this subsample in comparison to the total group. Roughly speaking, the coefficients for the two samples do not differ systematically (either in overall magnitude or in relation to changes in the schooltype-effect correlations).

Analyses of the influence of teacher survey variables on student achievement proceeded in two steps. The first set of analyses parallels the covariance analyses just described. Within a grade level cohort, analyses were conducted separately for data from the third and fourth years of the demonstration. In these analyses, the full set of teacher survey variables shown in Table 3.4 are entered as predictors of individual level student reading scores. These analyses are reported and discussed in Appendix C. Briefly, the third year results indicated no relationship between teacher variables and student outcomes. The fourth year results, although scattered across grade levels and levels of aggregation (i.e., class level and minischool level), suggested some interpretations worth examining further. The second set of analyses of teacher survey variables, reported here, follows up on these suggestions.

We wished to check for the possibility of separate classroom and program level effects of the teacher survey variables. To increase the sensitivity of the analysis to these effects, two steps were taken. First, student data were aggregated to the class level, after test scores were converted to a common metric,[7] so that all classes could

[7] The procedure for converting scores to a common metric relied on characteristics of the observed raw score distributions, as well as on the test publisher's conversion tables, since it was known that the conversions recommended by the test publisher, alone, were inappropriate for Alum Rock students. Within each grade level (test level and grade level are completely confounded), scores were standardized using the observed raw score mean and standard deviation, and then converted to the "extended standard score" metric using the appropriate mean and standard deviation from the published extended standard score distribution. To correct the analysis for possible abnormalities in the test score distributions across grade levels, the aggregate grade level for each class was also included as a covariate in the analysis.

be included in the same analysis. Second, preliminary analyses (i.e., those reported in Appendix Tables C.1 and C-2 and descriptive analyses of the aggregate data) were examined to identify a subset of the original 13 survey variables for inclusion in the final analysis. The results of this selection process appear in the variables entering the analyses presented in Table 3.5.

The results of the multilevel regression of class level reading achievement outcomes on the selected class and minischool level teacher survey variables are reported in Table 3.5 (coefficients for the background covariates have been omitted). The table contains the unstandardized regression coefficients (and t statistics in parentheses) for the teacher and minischool level variables. The variable, Principal Influence, is negatively keyed, so that a negative coefficient signifies increased student outcomes associated with an increase in perceived principal influence, and vice versa.

Perhaps the most noteworthy feature of the regression coefficients in the class and minischool level analyses is that the decomposition of effects into between- and within-minischool components results in opposite signed influences on student outcomes for each of the teacher survey variables. While the effects are not large (one coefficient is significant at the .05 level and two at the .10 level), the consistent patterns across the five variables and two aggregation levels overshadow this relatively weak showing. Staff Cohesion, Common Minischool Policies, and Principal Influence appear to be positively associated with student outcomes when viewed as properties of minischools, but are negative influences when viewed as characteristics teachers possess

Table 3.5

REGRESSION OF CLASS LEVEL STUDENT READING ACHIEVEMENT ON TEACHER AND MINISCHOOL LEVEL SURVEY VARIABLES[a,b]

Variable Name	Multilevel Regression		Class Level Regression
	Teacher Level	Minischool Level	
Staff Cohesion	-.225 (.487)	.679 (.849)	.197 (.536)
Common Minischool Policies	-.336 (1.445)	.360 (1.155)	-.057 (.371)
Teacher Autonomy	1.652 (1.601)	-2.256 (1.323)	.563 (.696)
Teacher Influence	.660 (1.714)	-1.330 (2.292)	-.135 (.508)
Principal Influence[c]	.545 (1.518)	-.961 (1.843)	.004 (.000)

[a]Unstandardized regression coefficients, $\underline{t}$ statistics in parentheses.

[b]The regression equation was estimated using the method of weighted least squares:

$$\underset{\sim}{\beta} = \{\underset{\sim}{W}'(\underset{\sim}{X}'\underset{\sim}{X})\underset{\sim}{W}\}^{-1}\ \underset{\sim}{W}'\underset{\sim}{X}'\underset{\sim}{y}\underset{\sim}{W} ,$$

where

$$\underset{\sim}{W} = tr(\underset{\sim}{n}_i)^{-1}k\underset{\sim}{n}_i = N^{-1}k\underset{\sim}{n}_i ;$$

$\underset{\sim}{\beta}$ is the vector of regression coefficients; $\underset{\sim}{X}$ is the matrix of independent variables; k is the number of classes; N is the total number of students; and $\underset{\sim}{n}_i$ is a diagonal matrix of class sizes. Use of the matrix $\underset{\sim}{W}$ insures that each classroom in the analysis will be weighted by the number of students contained in it, while the overall degrees of freedom for classes will be preserved.

[c]This variable was measured such that a negative coefficient represents greater influence.

relative to other teachers in their minischool. The reverse pattern is found for Teacher Autonomy and Teacher Influence. Students in minischools with high values on these variables have lower achievement, other things being equal, while teachers who are relatively more autonomous and perceive themselves to have greater influence than their minischool peers also have students with higher achievement.

The contrast of positive and negative effects at the two aggregation levels suggests a consistent interpretation for the results. Minischools with cohesive staff, common policies, and some degree of involvement from the school principal tend to demonstrate greater reading achievement than do less cohesive (etc.) minischools. Moreover, minischools consisting of teachers who, on average, conduct their work autonomously and exert considerable influence over decisionmaking tend to have lower achievement than minischools consisting of, on average, less influential and autonomous teachers. Since the minischool was the locus of decentralization of decisionmaking authority or, in other words, was the unit of implementation for the demonstration, it is reasonable to interpret effects at this level as related to the alternative school. Transferral of power to teachers is not associated with any absolute benefit to their students; minischools that are genuinely teacher run, with lower cohesiveness, shared policy, and a minimum amount of participation from the school principal (probably an appropriate form of treatment realization), are poorer environments for student learning. This does not, however, imply a generalized negative effect of the program, since minischools which functioned as integrated units with some supervision from the principal (also a legitimate

variant of treatment implementation) had mildly positive effects on students. Apparently, then, there was variety in the implementation of the program, with correspondingly varied effects on students.

For the effects at the teacher-within-minischool level, on the other hand, it is possible to construct a reasonable interpretation that is not directly tied to the implementation of the program at Alum Rock. Instead, effects at this level seem to be more indicative of personal and professional qualities of individual teachers, without regard to the extent the demonstration brought about minischool-wide changes (independence of minischool and teacher-within-minischool effects follows from the form of the analysis). For example, teachers within a given minischool who perceive themselves to have greater autonomy and influence than their peers (i.e., who have a high score on the within-minischool component of the Teacher Autonomy and Teacher Influence variables) might be teachers who possess a greater sense of efficacy regarding their work than teachers ranking low within their minischool on autonomy and influence. It is reasonable to suppose that more efficacious teachers are in fact better at the business of teaching (Berman and McLaughlin, 1977), and therefore to interpret the positive within-minischool effects of the autonomy and influence variables on students as related to a characteristic of teachers on which there would be individual differences whether or not there was high program implementation. Similarly, teachers whose perceptions of "closeknitness" (i.e., cohesion and common policies) were not shared by their minischool colleagues might lack the professional efficacy or autonomy needed for good teaching. By the same reasoning, teachers who

credit the principal with more control over their affairs than is commonly perceived in the minischool may also exert less control themselves. The negative within-minischool effects of Staff Cohesion, Common Policies, and Principal Influence are consistent with this interpretation.

The differing interpretations of between- and within-minischool effects are based on the rationale that the meaning of the teacher survey responses at the two levels is not constant across aggregation levels. Specifically, the minischool level variables are argued to be manifestations of the implementation of the educational program at Alum Rock. Hence, their impact on students is thought to be treatment related. The present data cannot establish, however, whether in the instances where the program appeared to negatively affect student performance, these effects came about because the program distracted teachers from instructional duties, raised their level of influence or efficacy to the point where conflicts began to surface, or disrupted school operations in some other manner. In any event, variations in aggregate teacher perceptions among minischools appear both to be program induced and to be associated with differences in the achievement of the students in those minischools. In contrast, relative effects within minischools seem to require a different conception of what the teacher survey responses measure. Apart from whatever impact the demonstration had on minischools, it is reasonable to expect teachers to differ from one another in attitude, orientation, and training, and for these differences to affect students. Thus, a conceptual analysis of the differences in meaning of the variables at the two levels of

aggregation leads to a framework for interpreting effects that provides a plausible account of the results.

The ability of the analysis to detect even weak effects of the teacher survey variables on students is noteworthy for two reasons: First, as conceived and implemented, the demonstration had no explicit objectives with regard to changing student performance; it was primarily an innovation in school finance and administration. Second, the variables included in the analysis are, at best, distal influences on learning and instruction; no direct indices of teachers' classroom practices were examined. Doubtless, a thorough explanation for the relationships observed in this study would need to involve an analysis of the classroom behavioral correlates and consequences of the organizational and attitudinal variables considered here.

Finally, it is instructive to compare the results of the analysis presented here with those obtained from the more conventional analysis entering teacher variables at the class level only. The results for this analysis are shown in Table 3.5. The pattern of results which ignore the minischool level bears little resemblance to the results from the multilevel analysis. In each case the coefficients for the teacher level are smaller than either of the corresponding coefficients from the multi-level regression. For two of the five variables, Common Policies and Teacher Autonomy, the sign of the total teacher level effect matches that of the within-minischool level effect; the total effect for Principal Influence is essentially zero; and for the remaining two variables, Staff Cohesion and Teacher Influence, the total effects appear to be more strongly influenced by the between-minischool level component from the multi-level analysis.

Program Size and Reading Achievement

An important aspect of the demonstration was the creation of smaller, semi-autonomous organization units (i.e., minischools) within participating schools in the district. Many changes in teacher attitudes and behaviors can be interpreted as responses to the establishment of these minischools. We wish to determine whether the experimental modifications of organizational size influenced student outcomes in ways not measured by our teacher survey variables. The data permit only an indirect examination of this issue. We can only test for residual program size effects, without the luxury of pinpointing the process variables responsible for such effects.

In the analyses, the program size variable is entered after removing the effects of background and teacher survey variables from the reading outcomes. Before reporting the results, however, several shortcomings should be mentioned. First, our measure of program size is a fallible one. The number of students in the program is the basis for our size estimate: We can only approximate this figure by counting those students in a program who appear on our data tape. Thus the range, or variability, in program size is likely to be restricted, since all inaccuracies will be errors of underestimation of actual size. Second, the design of the study does not allow us to distinguish between those residual effects due to program size and those due to other unmeasured, nonprogram-related peculiarities of the set of participating schools. The legitimacy of our interpretation, then, depends on our

having identified (and controlled for) at least the major nonprogram variables.

Table 3.6 presents the results of the program size analyses. In these analyses, four grade levels (cohorts) and two time spans are considered. Coefficients pertaining to the third and fourth years of the demonstration are presented.

The overall association between program size and reading achievement is reflected in the correlations shown in the table. Here, as well as elsewhere in the table, the effect of program size on student outcomes tends to fluctuate about zero during the third year of the demonstration, and to negatively covary with the fourth year outcomes. The pattern is the same both after background effects are removed and after background and teacher effects are removed (i.e., the standardized partial regression coefficients in the table). Larger programs, whether or not one takes account of the types of students entering them and the measured attitudes and perceptions of the teachers working in them, tend to inhibit achievement during the fourth year.

Student Classroom Perceptions and Reading Achievement

Are student perceptions of the level of positive affect, organization, ease of work, competitiveness, affiliativeness, and friction in their classrooms related to how well they perform on a test of reading achievement? To answer this question we consider data from the Classroom Environment Survey (CES) for the fourth year of the study. Our strategy in analyzing these data is to regress student reading achievement on the set of CES variables, along with the set of

Table 3.6

EFFECT OF PROGRAM SIZE ON READING ACHIEVEMENT

Grade and Years	Correlation[a] Size and Reading	Standardized Coefficients[b] for the Regression of Reading on Size Controlling for: Student Background	Student Background & Teacher Perception	N
Grade 3				
F'74-S'75	.090	.048	.040	383
S'75-S'76	-.060	-.040	-.133*	482
Grade 4				
F'74-S'75	.062*	.004	.008*	393
S'75-S'76	-.161*	-.041	-.124*	453
Grade 5				
F'74-S'75	-.014*	-.045*	.038*	443
S'75-S'76	-.136*	-.085*	-.101*	506
Grade 6				
F'74-S'75	.057	.040	-.006	450

[a]Correlation coefficients are starred if $p \leq .05$

[b]Standardized regression coefficients are starred if the corresponding unstandardized coefficients exceed twice their standard errors.

background control variables. We have chosen to disregard within-class variation in student perceptions. Consequently we employ CES data aggregated to the class level. This choice seems justified since disagreements among students within classes more likely reflect individual biases than true differences in the classroom environment itself.

Table 3.7 presents the results of the analysis of the effects of student classroom perceptions on reading achievement outcomes.[8] In general, student perceptions of classroom environment seem unrelated to

Table 3.7

EFFECTS OF STUDENT CLASSROOM PERCEPTIONS ON READING ACHIEVEMENT IN THE FOURTH YEAR OF THE DEMONSTRATION: STANDARDIZED REGRESSION COEFFICIENTS WITH COVARIATE ADJUSTMENT

Variable	Grade Level		
	3	4	5
Liking[a]	-.030	-.109	.087
Organization	.062	.060	.087
Ease	.006	-.126*[b]	-.136*
Competition	.008	.049	-.108*
Friendliness	.006	-.061	.076*
Friction	.046	.000	-.040
N	305	300	358

[a]Variables are defined in Section II.

[b]Standardized regression coefficients are starred if the corresponding unstandardized coefficients exceed twice their standard errors.

[8] It should be noted that because of the small size of the samples which were administered the CES, we have combined participating and nonparticipating students in the same analysis.

reading achievement outcomes. The major exception to this conclusion is the variable Ease of Work. The negative relation between Ease and Achievement in the fourth and fifth grades (-.126 and -.136, respectively) indicate that as the difficulty of classwork increases, so does achievement, holding constant all other variables. This finding is consistent with findings from the BTES reported by Burstein (1980).

The Effects of Program Choice and Nonlocal Attendance on Reading Achievement

One central feature in voucher theory and in the set of innovations introduced in Alum Rock is the option for students to enter and leave alternative-education programs. Programs offering "popular" education will have relatively more students transferring into them than will unpopular ones; specialized programs will attract pupils with congruent needs or interests. The system of alternatives also removed geographical restrictions on the choice of educational programs. Students and parents were free to choose from among all participating schools, without the usual expense associated with choosing a school outside their "catchment area."

Program choice can be viewed both as a factor regulating the growth and decline of minischools, and as a characteristic of individual students. With respect to individual students the policy question is not whether the educational system responds according to free market principles, but rather whether students benefit from making educational choices.

Student choice data in aggregate form provide information about the classroom and program environments in demonstration schools. Variation

from program to program (or class to class) in enrollment stability as reflected in the aggregated student mobility figures may relate to school learning and performance quite differently than would the disaggregated history of program choices of an individual pupil. Programs with high student turnover rates may actually depress performance. An analysis of the effects of student choice on achievement, then, should attempt to separate such effects into the components to which distinct interpretations may be attached.

In analyzing and interpreting the effects of program choice, we would like to know whether choices were made on the basis of interest in the chosen program, or information about it. Unfortunately, such data were not available at the time our analyses were conducted. Because of this, our analyses provide only indirect evidence on the effects of this demonstration feature. Two types of available information pertain to student choice:

1. Program changes-- For each student we counted the number of times a change in program attendance occurred. Data collection involved up to nine contacts with each student, and for each of these occasions we could determine whether the student was in the same or a different program than the previous occasion.
2. Non-local attendance-- The program choice option allowed students to choose non-local schools. For each student we counted the number of data collection occasions at which a non-local school was attended. We consider non-local school attendance important because it removes some of the ambiguity about program choices: Children presumably attend non-local schools for reasons other than apathy or convenience.

The analyses focus on the effects of these two variables on student reading achievement after adjusting for the effects of student background. Controlling for student family background is especially important in these analyses because of the likely relation between SES and mobility. The variable choice in each analysis is split into two components, one indicating frequency of occurrence prior to fall 1974 and the other indicating frequency of occurrence from fall 1974 through spring 1976. The effect of each component of the choice variables is assessed separately (in multilevel fashion for the fall 1974 through spring 1976 component), while holding the other constant. Since students enrolled in 3rd-6th grades in the fourth demonstration year will have had differing amounts of experience in alternatives prior to fall 1974 (i.e., during the first two years), we have included as an additional covariate the number of testing periods during this interval each student was enrolled in an alternative school.

The results of the program choice and non-local attendance regression analyses are reported in Table 3.8. The two components of each independent variable are shown separately. In the portion of student movement taking place before our background control variables are measured, we see the effects of early program choices and nonlocal attendance, holding constant background and later choices. The "total" effect of later program choice and nonlocal attendance is also shown, including (1) the effect on achievement for students who move relatively more often than their classmates; (2) the effect for classes with more aggregate movement than other classes in the same minischool program;

Table 3.8

EFFECTS OF PROGRAM CHOICE AND NONLOCAL SCHOOL ATTENDANCE ON READING ACHIEVEMENT: STANDARDIZED PARTIAL REGRESSION COEFFICIENTS WITH COVARIATE ADJUSTMENT

Analyses \ Grade Level	Program Choice				Nonlocal Attendance			
	3	4	5	6	3	4	5	6
A. Total Individual Level								
1. Early choice controlling for later choice	-.052	-.027	.009	-.025	.075	.027	-.133*	-.061*
2. Later choice controlling for early choice	.026	-.029	-.053	-.020	-.029	-.004	.085	.071
B. Multilevel								
1. Individuals within classes	.026	-.034	-.050	-.042	-.026	.005	.094*	.016
2. Classes within minischools	.007	.040	-.019	.044	.024	-.085*	-.017	-.027
3. Between minischools	-.015	.042	.004	.017	-.126*	.046	-.070	.022
N	425	436	463	606	425	436	463	606

*Standardized regression coefficients are starred if the corresponding unstandardized coefficients exceed twice their standard errors.

and (3) the effect for programs with higher aggregate mobility than other programs. All coefficients for program choice are small and fail to reach statistical (or practical) significance. Over and above the influence of students' background and prior performance, the frequency with which they make different program choices contributes nothing to the prediction of their reading achievement.

With respect to nonlocal attendance, throughout the demonstration, the analysis suggests a weak negative relation to achievement. The data are by no means consistent on this point, however; several small positive coefficients (one statistically significant) can be found in Table 3.8.

IV. SUMMARY AND IMPLICATIONS OF THE EFFECTS OF ALTERNATIVE-SCHOOL PROGRAMS[1]

VOUCHERS, ALTERNATIVE-SCHOOL PROGRAMS IN ALUM ROCK, AND STUDENT OUTCOMES

The concept of an education voucher is compellingly simple and seductive to those who seek policy levers to motivate educational improvement. Rather than give money to public schools, parents would receive a voucher equal in value to the per pupil expenditure for their local public school. They could spend this voucher to educate their children in the public, private, or parochial school of their choice. Parental choice, a cornerstone of the voucher concept, would create a "free-market education" economy. Competition would loosen up the public school system and lead to the creation and survival of "popular" and, in the minds of some policymakers, "high quality" educational programs.

In breaking the public schools' monopoly on government-financed education, vouchers were expected to lead to a variety of alternative-education programs. The programs were expected to evolve in response to differences in parents' educational choices for their children; these choices, presumably, would reflect their children's needs. Educational programs, then, were expected to be especially well suited to their students. Without the monopoly and accompanying burdensome administrative structure of the public schools, the decisionmaking power

[1] This section was written by Richard Shavelson and Frank Capell.

of teachers and school principals was expected to increase; increased local decisionmaking was expected to lead to educational programs that were responsive to students; and parents were expected to directly affect schools rather than have their influence diffused. Finally, as a potential by-product of vouchers, students were expected to perceive their schools, themselves, and their peers more positively and to achieve at higher levels because their needs, interests, and capabilities were better matched to their educational experiences.

For a wide variety of reasons, the program implemented in Alum Rock was not a true voucher program. Parents did not have the choice of private education since there were no private schools in the area due to the low economic status of the families. Even if there had been, California law (though changed during the second year of the demonstration) prohibited the use of state funds for private or parochial schools. Parents' choice of programs through the use of vouchers, therefore, was limited to selecting one of several programs initiated within the existing public schools. Each participating school in Alum Rock offered two or more "minischools," alternative schools within the school.

The competitive benefits of a true voucher program were somewhat lost since teachers were guaranteed their salaries even if their ternative programs had low enrollments or ceased to exist. Moreoever, ministrators and teachers insisted that minischool enrollments be stricted; and, at the same time, parents wanted more than one program d the right to priority enrollment in their own neighborhood school.

Alum Rock, then, did not offer the free-market education underlying the concept of an education voucher. Politics won out over economics; innovations are always adapted to the local setting. At the height of the demonstration, what Alum Rock offered was 50 minischools, all purporting to offer alternative forms of education. Nevertheless, the demonstration was extremely informative about some aspects of vouchers because, like the voucher concept, it offered (in diminished form) the following: (a) parental choice, (b) alternative-education programs in the form of minischools, (c) decentralized decisionmaking with teachers and principals receiving increased power to make curricular choices and budget allocations, and (d) smaller schools.

The purpose of this report was to examine the effect of the demonstration on students' achievement (cognitive outcomes) and perceptions of their schools, themselves, and their peers (noncognitive outcomes). In summarizing and drawing implications of the results of the study, we emphasize the fact that vouchers, and the Alum Rock demonstration in particular, were directed toward the goal of greater choice of alternative, responsive forms of education. Student cognitive and noncognitive outcomes would be no more than a potential by-product of achieving this goal. Vouchers and the demonstration should not be evaluated solely on the basis of cognitive and noncognitive outcomes. Rather, the findings presented here should be used to improve the implementation of educational alternatives and the quality of the alternative education programs offered.

FINDINGS IN REVIEW

This study sought answers to two broad questions: Do alternative-school programs affect student achievement and perceptions differently than do traditional-school programs?" and "What features of alternative-school programs (if any) are associated with increased achievement?" The answers to these questions, however, are inevitably incomplete. Traditional- and alternative-school programs espouse multiple, differing goals, both about the processes of education and the effects of these diverse processes on students. A comprehensive assessment of the impact of these programs, then, is virtually impossible to achieve with a reasonable number of measures in a reasonable amount of time. Moreover, the very nature of alternative-school programs with teacher-school choice in participation and parent-student choice in selection of a program results in a quasi-experimental rather than a randomized-experimental design. As a consequence, the data on student outcomes had to be adjusted statistically for differences in students attending traditional- and alternative-school programs. These adjustments always leave room for some degree of ambiguity in interpreting the effects of alternative-school programs on students' achievement and affect. Finally, alternative-school programs have been implemented in something less than their ideal form for three years' duration at most. Due to restrictions in implementation and time, any estimate of the effect of such programs on student outcomes must be incomplete. With these caveats in mind, the results of the study are summarized below.

Effect of Alternative- and Traditional-School Programs on Student Outcomes

Student Achievement in Reading. Reading achievement test scores were available for students in traditional- and alternative-school programs in grades 3 through 6 who had participated in their respective programs for 1, 2, or 3 years. In examining the effect of traditional- and alternative-school programs on reading, posttest scores were first adjusted for differences between students in reading pretest scores, ethnicity, socioeconomic status, and whether Spanish was spoken in their homes. Then the effect of traditional- and alternative-schools on students' (adjusted) reading posttest scores was ascertained.

The findings can be summarized succinctly. No appreciable or consistent differences in students' adjusted reading achievement were found between traditional- and alternative-schools. This finding held for students in different grade levels and for students participating in traditional and alternative programs for differing lengths of time.

Student Noncognitive Outcomes. We examined the effects of traditional- and alternative-school programs on students' perceptions of social distance from significant others and their perceptions of themselves (self-esteem) and their peers. The data on students' perceptions were gathered in the third and fourth years of the demonstration in grades 3 through 6. In examining the effect of traditional- and alternative-schools on each perception variable, posttest scores on, for example, self-esteem were first adjusted for differences between students in perceptions at pretest (i.e., scores on self-esteem at pretest), ethnicity, socioeconomic status, and whether

Spanish was spoken at home. Then the effects of traditional- and alternative-schools on students' adjusted posttest perceptions were ascertained. No appreciable or consistent differences in students' self-esteem or in their perception of peers were found between traditional- and alternative-schools. Nor were there differences in self-esteem and peer perception between grade levels or between certain grade-level/school type (traditional or alternative) combinations.

However, in comparision to students in traditional schools, students in alternative schools perceived themselves to be slightly more distant from significant others such as teachers. Furthermore, while the tendency for students in traditional schools is to perceive social distance as decreasing in later grades, students in alternative schools perceived greater social distance in higher grades. However, without a clear account of the factors combining to produce differences in students' standing on the pretest of social distance, several alternative explanations for these findings may be set forth: (a) alternative schools, on average, had a slightly negative effect on students' perception of social distance from significant others, (b) alternative schools, on average, slightly increased the students' perceived social distance in the fourth year after the first three years in which perceived social distance was smaller, or (c) bias in statistical adjustments rather than a slight increase in social distance accounts for the findings.

Effects of Various Features of Alternative Schools on Student Reading Achievement

Sometimes educational policies are implemented with the goal of improving education without even hinting at the processes that may, with some probability, bring about educational improvement. Education vouchers are an example of such policy. Schools, in response to "free-market" competition, are expected to provide improved educational alternatives. However, voucher policy does not mention anything about how this is to be accomplished. One important contribution of our study of alternative-schools in Alum Rock, then, is to identify features of these schools which are associated with student achievement. This information might aid other schools in improving education, putting alternatives into effect, or both. In this section we summarize the findings on the effects of certain features of alternative-schools on students' reading achievement. More specifically, we examine the effects on reading achievement of teachers', and students' perceptions of their alternative schools, program size, and parent choice.

Teachers' Perceptions and Reading Achievement. Information about teachers' perceptions of staff cohesion, shared minischool policy, their autonomy and influence, and their principal's influence was available for the third and fourth years of the demonstration. The effect of each of these perception variables on student reading achievement was examined after reading achievement was adjusted for the other teacher

perceptions, as well as student reading pretest scores, ethnicity, SES, and whether Spanish was spoken at home.[2]

Before we report the findings, however, we need to point out that we estimated the effect of each teacher perception variable on adjusted student achievement, both at the level of the classroom and at the level of the minischool (i.e., data were aggregated over classrooms within a minischool). We interpret the effects at the minischool level to be associated primarily with the implementation of the alternative educational program. Hence, the effect of teachers' perceptions on student achievement, examined at the minischool level, bear on programmatic implementation of decentralized decisionmaking. Since perceptions of cohesiveness, common policy, and principal's influence reflect the program as a whole, they might reasonably be expected to have a positive influence on achievement at this level of aggregation. In contrast, we view the effects at the level of the classroom to reflect individual differences between teachers within a program (minischool). That is, teachers may vary in their willingness and ability to achieve the goals of the alternative program in their classrooms. Those teachers who perceive themselves more autonomous and influential than their colleagues in the minischools might be associated with individual classrooms which tend toward higher achievement (see Chapter III).

We found that minischools in which teachers, on average, perceived a cohesive staff, common policies, and some degree of involvement from the principal were associated with higher reading achievement than minischools with teachers perceiving a lack in cohesion, common

[2] Third year results are reported in Appendix D.

policies, and principal involvement. However, minischools in which teachers, on average, perceived greater autonomy and influence were associated with lower reading achievement.

Within a typical minischool, we found that classrooms of teachers who perceived themselves as having greater autonomy and individual influence were associated with higher reading achievement than were classrooms of their colleagues who perceived themselves as having less autonomy and influence. However, within a minischool, teachers who viewed their alternative-education program as having greater cohesiveness, common policies, and principal involvement than their colleagues tended to be associated with lower achieving students. We interpret this finding to mean that teachers who perceive themselves as autonomous and influential are associated with increased student achievement, regardless of minischool affiliation. And teachers whose perceptions of "closeknitness" were not shared by their minischool colleagues might not, individually, exert the needed classroom leadership associated with achievement. In short, cohesive minischool programs comprised of teachers who did not perceive themselves as especially autonomous or influential were associated with higher student achievement in reading while individual teachers who perceived themselves as autonomous and influential were associated with higher achievement, regardless of the cohesiveness of their minischool programs.

Students' Perceptions of Classroom Environment and Their Reading Achievement. Information about students' perceptions of the level of positive affect, organization, ease of work, affiliativeness, and

friction in their classrooms was available for students in grades 3 through 5 during the fourth year of the study. The effects of each of these perceptions (e.g., ease) on student's reading achievement were examined after achievement was adjusted for the relevant pretest (i.e., ease) and background variables (e.g., ethnicity, SES). In these analyses, we chose to disregard within-class variation in student perception and aggregated the perception data to the class level. (Disagreements among students within classes more likely reflected individual biases than true differences in the classroom environment.)

No appreciable or consistent effects on student reading achievement were found for students' perceptions of the level of positive affect for the class, the organization of the class, the competition in the class, the friendliness of the class, or the friction in the class. However, the ease of work was related to student achievement: as the perceived difficulty of classroom work increased, so did mean scores on reading achievement in grades 4 and 5.

Program Size and Reading Achievement. The demonstration, in creating minischools, created semiautonomous organizational units. Minischools were smaller than traditional schools. Because of their smaller size, minischools might be expected to facilitate communication among students, teachers, administrators, and parents which might, in turn, lead to improved student achievement. While we do not have a measure of this communication flow, we can examine the effect of program size on student achievement for students in grades 3 through 6 during the third and fourth years of the demonstration. In analyzing the data, the effects of student background characteristics (e.g., ethnicity, SES)

and teacher perception and attitude data (e.g., shared policy, independence) were removed from the reading achievement scores. Then the effect of program size on the adjusted reading scores was estimated.

In the third year of the demonstration, the relation of program size to student reading achievement was not reliably different from zero. There was, however, a slight consistent (not statistically significant) tendency for larger program size to be associated with higher achievement. In contrast, larger program size was associated with lower reading achievement in the fourth year of the demonstration.

Program Choice, Non-Local Attendance and Reading Achievement. One of the central features of the demonstration (and of vouchers) is the opportunity for parents to choose where their children will go to school. By providing parental choice, presumably, a closer match will be achieved between students' abilities and interests and the education they receive. Moreoever, minischools offering "popular" education programs attract students, while "unpopular" programs are not competitive and so are either changed or dropped. Further, the demonstration (and vouchers in general) did not limit parental choice geographically. Parents could choose any public alternative school in the district.

Two types of information were available for examining the effect of parental choice on reading achievement: (1) the number of program changes made by a student during a school year and (2) the number of data collection periods (i.e., the fall and spring of each year) a student attended a nonlocal minischool. Both types of information serve as proxies for data on parent choice such as whether a conscious choice was made and what the reasons were for the choice.

In examining the effect of the number of changes on student achievement, the effects of student background (e.g., reading pretest scores, ethnicity, SES), and amount of early alternative school experience were first removed from the reading scores. Then the relation between changes and (adjusted) reading scores was estimated. The effects of nonlocal attendance on reading scores was estimated in a similar manner, first removing the effects of number of program changes rather than nonlocal attendance from reading scores.

No appreciable or consistent effects on reading achievement were found for the number of program changes made by a student or the number of occasions a student attended a nonlocal minischool. Parental choice, at least as measured by these two proxy variables, appears to be unrelated to student achievement in reading.

IMPLICATIONS OF THE EFFECTS OF ALTERNATIVE-SCHOOL PROGRAMS ON STUDENT OUTCOMES

Education vouchers provide policymakers with an economic policy lever for motivating educational change. However, the voucher concept is insensitive to the political realities of schools and, as a consequence, does not provide mechanisms for implementing alternative-education programs. Moreover, vouchers are insensitive to the processes of education and therefore do not identify such processes and their probable outcomes.

While limited in scope, our study of the effects of education alternatives on student outcomes in Alum Rock suggests certain

mechanisms for implementing alternatives and certain perceptions of teachers that may lead to increased student achievement. Though limited, these findings provide an initial step toward informing those who would implement education vouchers or alternative schools as to certain things that could be done to enhance student (reading) achievement.

Alternative-education programs in Alum Rock neither hindered nor enhanced student achievement and affect, compared to traditional-education programs. That the alternatives did not hinder student outcomes is an important finding since most of the programs had less than a three year history and so were only partially implemented. One might reasonably expect partially implemented programs not to fare as well as long-established programs with which teachers, students, and parents are familiar.[3] This leads us to recommend that

> the justification for implementing alternative-education programs probably should not be based on the promise of enhanced student outcomes. Rather, the justification of alternatives might be based on public policy (e.g., education vouchers) or public pressure (demand for greater parent choice). Conversely, the argument against education vouchers or other policies which might reasonably lead to alternative-education programs probably should not be based on the fear of hindering student achievement or affect.

[3] Perhaps one possible explanation as to why the implementation process did not hinder student outcomes is that 48 of the 51 alternative programs stressed basis skills, in spite of their publicized differences, and so looked very much like their traditional counterparts (Barker, Bikson, and Kimbrough, 1981).

The process of implementing an innovative educational program is crucial to its success and survival. Berman and McLaughlin (1978) cited two keys to success and survival: (1) shared support of the program by teachers and administrators, especially principals, and (2) a plan for incorporating the innovation into the school at the outset. Our findings of the effects of teachers' perceptions on student outcomes bear on the shared support aspect of implementation. Alternative-education programs perceived by participating teachers as cohesive with a common policy and principal support were associated with higher achievement than alternative programs characterized by less cooperation and greater perceived teacher autonomy and influence. This leads to the recommendation that

> in implementing alternative-education programs in response to public policy (e.g., vouchers) or pressure (e.g., parental choice), careful attention should be given to developing teacher and principal support for the program. To the extent that the alternative is successfully implemented in this manner, students may benefit from these programs over programs implemented with little perceived teacher and principal support.

However, the Alum Rock findings also indicate that classrooms of teachers who perceived themselves as more autonomous and influential than other teachers were associated with higher achievement. We interpret this finding to mean that teachers differ from one another and

that certain teachers will have higher achieving classes than other classes, regardless of the quality of the alternative-education program. This leads to the recommendation that

> in the selection of individual teachers attention should be paid to possible tradeoffs between autonomy and cohesiveness. Selection of autonomous, influential teachers may lead to noteworthy alternative-education classrooms, at least as measured by student outcomes, even if at the expense of a cohesive overall program. [4]

Finally, a cornerstone of vouchers and of the demonstration is parental choice. We found that parental choice was unrelated to student (reading) achievement. There are several reasons why we consider this finding tentative. First, we did not have actual choice information on which to base the analysis. Rather, we used proxies such as the number of programs a student attended. Clearly, students may change education programs for a wide variety of reasons, only some of which might be related to achievement. Perhaps just as important was the fact that most parents wanted to send their children to neighborhood schools. That is, parental choice may be based on factors other than the best match between child and education program. Proximity is probably one such factor. It just may be that vouchers or other policies directed at parental choice as a vehicle for motivating educational change may be naive with respect to factors motivating parental choice itself. This

---

[4] The recommendation is consistent with the findings of Barker et al. (1981)

is an important issue in need of further research before policy recommendations can be made.

Appendix A

CORRELATIONS

Table A.1

CORRELATIONS OF MAT TOTAL READING, TOTAL MATH AND SUBTOTAL MATH SCORES

Cohort	Grade	Test Administration	Test Level	Total Reading with Total Math[a]	Total Reading with Subtotal Math[b]	Total Math with Subtotal Math
1	2	Fall 1974	Primary I	.622	*[c]	*
		Spring 1975	Primary II	(.692)[d]	.665	(.999)
	3	Fall 1975	Primary II	.685	.672	.973
		Spring 1976	Primary II	.707	.696	.976
	4	Fall 1976	Primary II	.707	.673	.980
2	3	Fall 1974	Primary I	.642	*	*
		Spring 1975	Primary II	(.725)	.704	(.999)
	4	Fall 1975	Primary II	.735	.706	.975
		Spring 1976	Elementary	.716	.686	.981
	5	Fall 1976	Elementary	.762	.725	.982
3	4	Fall 1974	Primary II	**[e]	.700	**
		Spring 1975	Elementary	(.714)	.699	(.999)
	5	Fall 1975	Elementary	.750	.717	.980
		Spring 1976	Elementary	.758	.727	.982
	6	Fall 1976	Intermediate	.686	.647	.975
4	5	Fall 1974	Elementary	**	.707	**
		Spring 1975	Elementary	(.742)	.728	(.999)
	6	Fall 1975	Intermediate	.633	.633	.977
		Spring 1976	Intermediate	.685	.651	.982
	7	Fall 1976	Intermediate	.720	.681	.981

[a]"Total math" = sum of 3 subtests except in Spring 1975 when a total was calculated from 2 subtests using a formula derived by Rand analysts.

[b]"Subtotal math" = sum of 2 subtests (computation and concepts).

[c]The asterisk indicates that, for this test level, only one subtest is given and is recorded at the total score.

[d]Parenthesized coefficients involve the "constructed" total score.

[e]The double asterisk indicates that no total math was reported for this administration.

Appendix B

MEANS, STANDARD DEVIATIONS, AND SAMPLE SIZES

Table B.1

TOTAL SAMPLE STANDARD SCORE MEANS, STANDARD DEVIATIONS AND SAMPLE SIZES FOR STUDENTS IN ALTERNATIVE (A) AND REGULAR (R) SCHOOLS

	Fall 1974		Spring 1975		Fall 1975		Spring 1976		Fall 1976	
	A	R	A	R	A	R	A	R	A	R
1st Grade										
Mean			31.82	34.53	25.32	27.35	36.83	37.26		
SD			9.42	10.47	5.73	5.60	9.72	9.10		
N			673	583	758	650	765	516		
2nd Grade										
Mean	36.73	38.84	46.04	48.11	35.64	39.91	47.70	49.37	35.36	40.32
SD	10.12	9.90	10.77	11.29	10.02	10.92	10.85	11.03	10.47	10.66
N	661	552	737	607	722	532	763	504	678	625
3rd Grade										
Mean	46.69	49.64	53.88	56.25	48.40	50.84	56.59	58.76	46.38	50.54
SD	10.43	10.53	9.90	10.25	10.93	11.31	10.53	11.99	10.44	10.51
N	712	511	739	559	782	590	798	492	698	597
4th Grade										
Mean	55.84	58.54	58.86	62.30	56.5	58.25	61.24	62.08	52.31	57.02
SD	10.62	10.09	10.63	11.76	10.84	10.62	11.65	11.66	13.15	11.95
N	727	580	777	622	764	566	758	485	700	598
5th Grade										
Mean	60.38	62.21	63.49	68.05	60.47	53.57	66.35	69.19	59.69	63.11
SD	12.74	12.67	11.57	12.33	12.36	13.13	12.45	12.89	11.75	12.06
N	773	593	785	608	802	608	814	508	732	624
6th Grade										
	65.66	73.41	70.85	79.36	66.38	70.28	72.45	76.80	67.05	70.12
	13.73	14.53	13.62	14.24	13.09	12.66	12.27	11.90	12.42	12.40
	940	352	934	379	1014	374	1018	319	945	422
de										
	72.63	80.01	76.61	83.65	72.60	81.03	78.75	83.56	72.44	76.39
	13.83	13.68	112.79	14.75	13.17	15.13	14.00	13.19	12.27	11.26
	940	354	947	339	973	358	981	304	988	392
de										
	79.70	83.77	83.35	90.32	77.21	85.53	80.69	90.43	75.44	79.75
	14.91	15.66	14.64	15.59	13.67	14.35	14.71	13.16	14.23	14.53
	993	319	1027	315	961	371	972	326	992	354

Table B.2

ANALYSIS SAMPLE STANDARD SCORE MEANS, STANDARD DEVIATIONS AND SAMPLE SIZES FOR ALTERNATIVE (A) AND REGULAR (R) STUDENTS

	Fall 1974		Spring 1975		Fall 1975		Spring 1976		Fall 1976	
	A	R	A	R	A	R	A	R	A	R
1st Grade										
Mean			32.46	35.50						
SD			9.40	10.47						
N			486	404						
2nd Grade										
Mean	36.50	39.13	45.88	47.82	35.94	40.59	47.7	49.37		
SD	10.15	9.06	10.72	10.89	10.15	10.94	10.88	11.03		
N	481	351	560	418	636	458	763	504		
3rd Grade										
Mean	46.75	50.43	53.77	56.34	48.69	51.14	56.59	58.76	46.67	50.63
SD	9.91	10.31	9.53	9.62	10.88	11.21	10.53	11.99	10.63	10.41
N	510	348	576	403	679	499	798	492	575	458
4th Grade										
Mean	55.73	58.67	59.12	61.92	56.81	58.70	61.24	62.08	52.98	57.19
SD	9.93	9.96	10.31	11.35	10.47	10.70	11.65	11.66	13.05	11.52
N	545	399	626	450	673	491	758	485	576	472
5th Grade										
Mean	59.33	61.66	64.04	67.37	60.77	64.05	66.35	69.19	60.19	63.73
SD	11.53	11.42	11.19	10.50	12.18	12.76	12.45	12.89	11.68	11.93
N	562	411	608	445	729	537	814	508	579	501
6th Grade										
Mean					66.57	70.47	72.45	76.80	67.29	70.27
SD					13.07	12.37	12.27	11.90	12.28	12.37
N					909	337	1018	319	763	351
7th Grade										
Mean									72.35	76.44
SD									11.66	11.16
N									809	301
8th Grade										
Mean										
SD										
N										

Appendix C

READING SCORE DISTRIBUTIONS

There has been considerable debate over the assignment of appropriate test levels to children, in particular over the effects incorrect level assignnment can have on test score characteristics (e.g., "floor" and "ceiling" effects). We present here the frequency distributions of scores for the test levels used in our analyses and the administration on which reliability estimates were based.

Figures C.1 through C.4 show the pair of distributions for each test level. The distributions appear to be only moderately skewed, and inconsistent in the direction of skew (skewness ranged from .575 to -.231). A more marked feature of the distributions for the first two levels is the rather clear bimodality, especially in the fall. One possible explanation of this characteristic of the distributions is that they represent a combination of scores from two populations with different means. Figures C.5 and C.6 present the distributions for these test levels separately for students in alternative and regular schools. As can be seen, the bimodality remains in each distribution. Hence, an alternative explanation must be sought.

Since the scores represented in the figures are totals over a number of subtests, it may be that the interrelations among subtests vary as a function of period of measurement (i.e., fall vs. spring), perhaps interacting with student characteristics. For example, differences in the environments students experience over the summer may affect the subtests differently. One implication of this argument would

be that the dispute over the determination of appropriate test levels for children might need to be focused on the levels of difficulty of subtests rather than on total tests, especially for students at earlier grade levels. We have not systematically explored these hypotheses, so at present they remain as important areas for future analyses of the Alum Rock data.

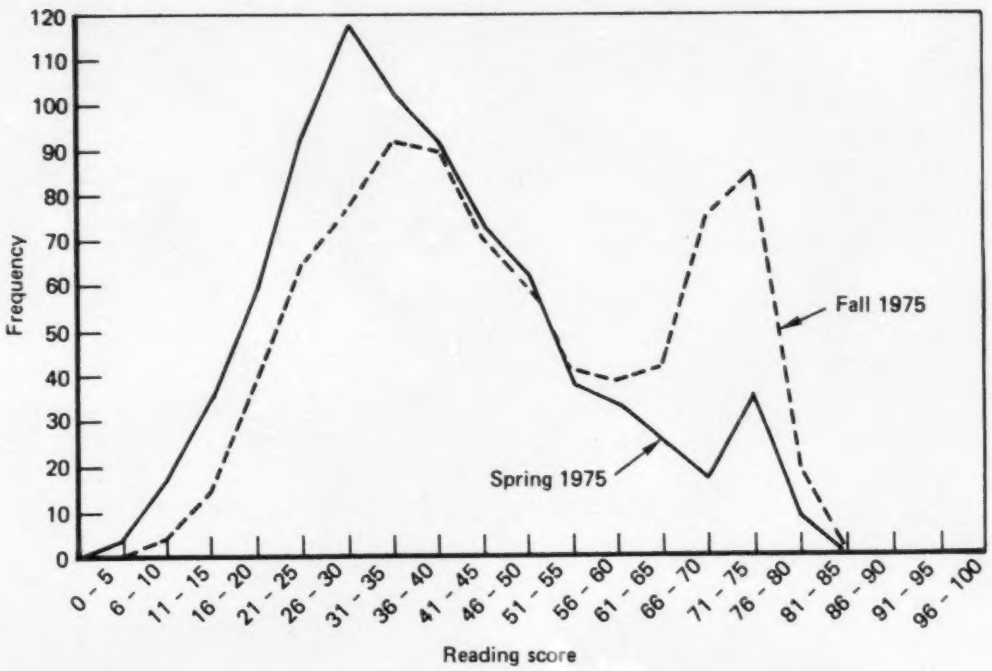


Fig. C.1 — Reading scores - 2nd graders, spring 1975, and 3rd graders, fall 1975

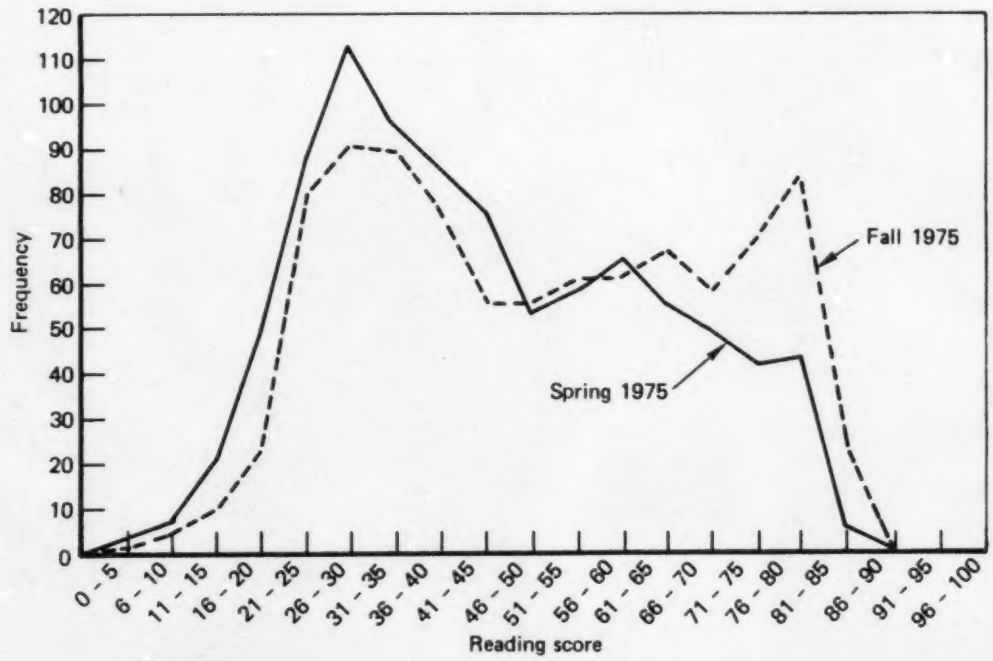


Fig. C.2 — Reading scores - 3rd graders, spring 1975, and 4th graders, fall 1975

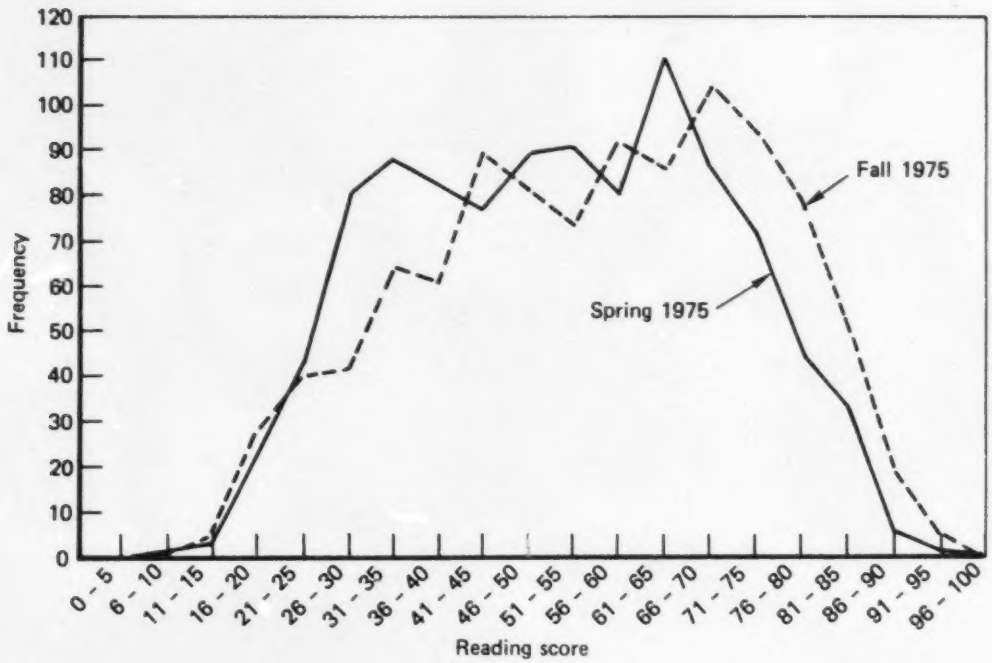


Fig. C.3 — Reading scores - 4th graders, spring 1975, and 5th graders, fall 1975

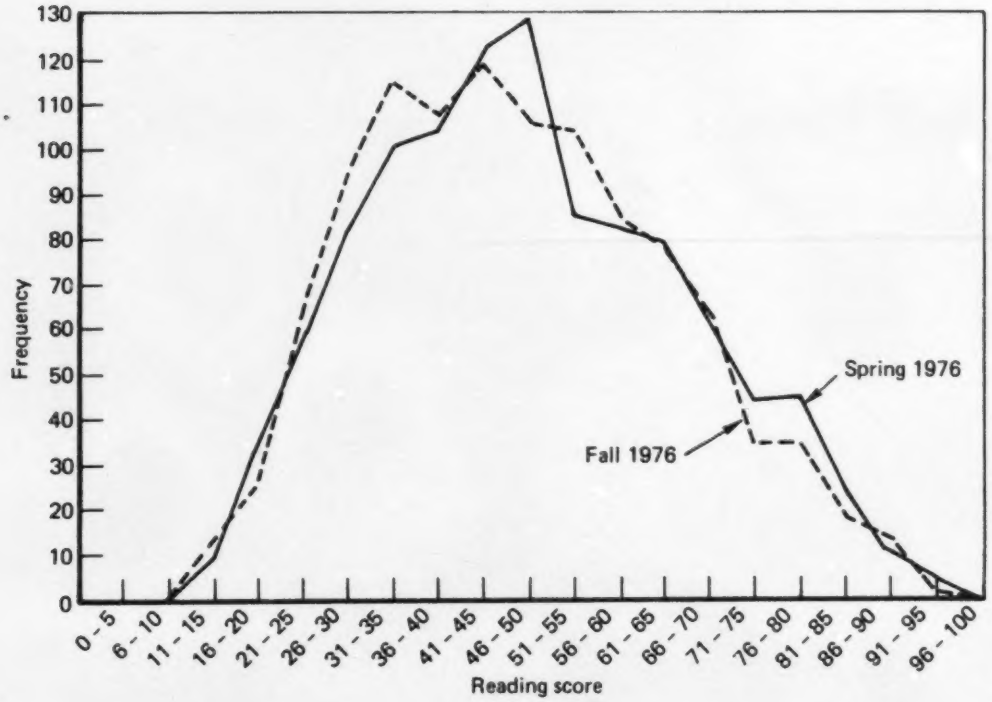
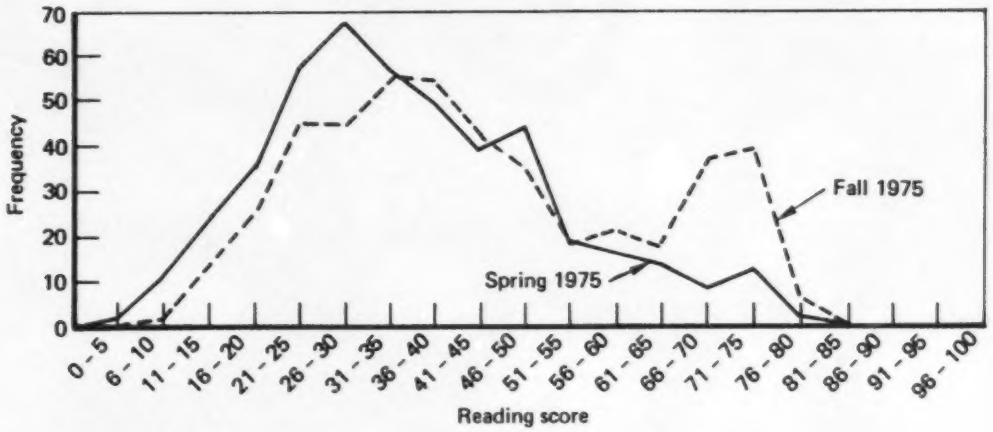
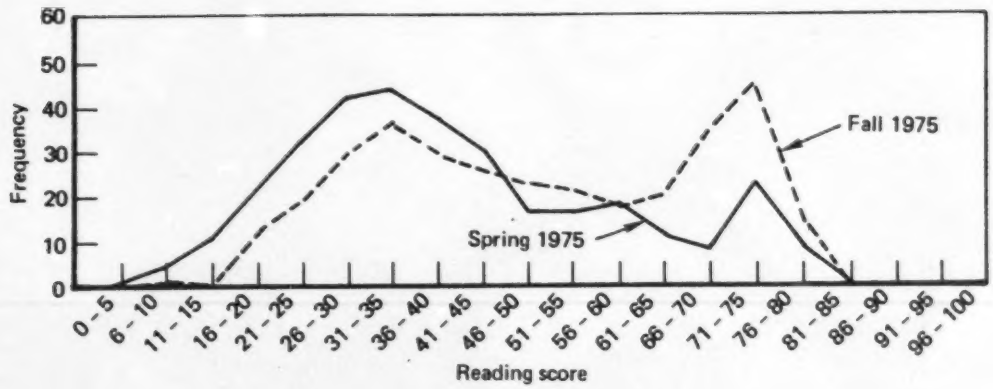


Fig. C.4 — Reading scores - 6th graders, spring 1976, and 7th graders, fall 1976

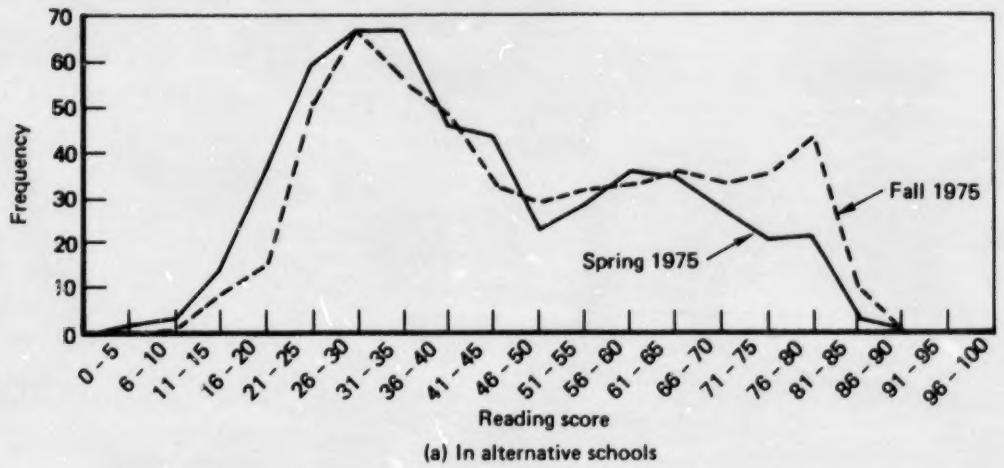


(a) In alternative schools

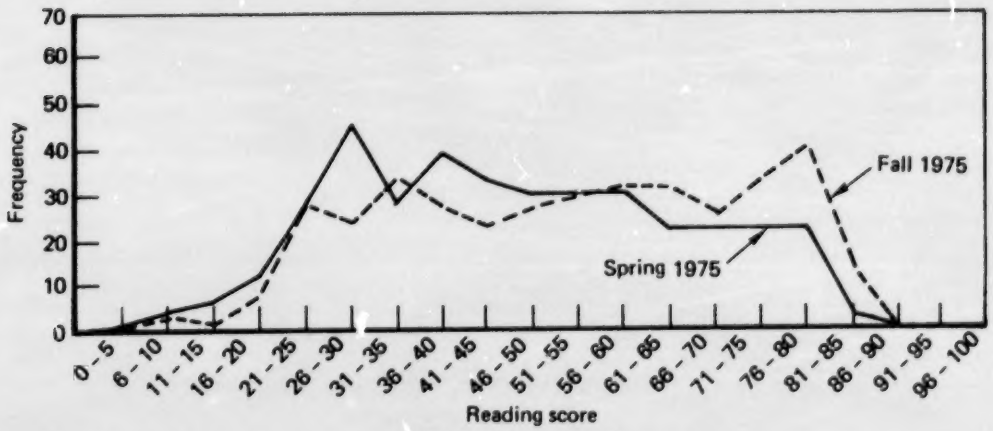


(b) In regular schools

Fig. C.5 — Reading scores - 2nd graders, spring 1975, and 3rd graders, fall 1975, alternative and regular schools



(a) In alternative schools



(b) In regular schools

Fig. C.6 — Reading scores - 3rd graders, spring 1975, and 4th graders, fall 1975, alternative and regular schools

Appendix D

PRELIMINARY ANALYSES OF THE FULL SET OF TEACHER SURVEY VARIABLES

Tables D.1 and D.2 present the results of four sets of regression analyses of the full set of teacher survey variables.

Table D.1 contains analyses predicting end of third year reading achievement from third year teacher survey responses entered both as individual teacher scores and as minischool means of teacher scores. The analyses include only teachers and students in participating programs. Control variables, not included in the table, are the standard set from previous analyses. The first feature to note about the table is the nearly complete absence of any significant effects. Of the 104 coefficients, five reach statistical significance; this result is the sampling expectation in the case where the true value of every coefficient is 0. There appears no reasonable choice but to report that no findings of any practical importance are suggested by the table.

One possible explanation for this result is that the control variable set accounts for all the predictable variance in the outcome measures: R^2 (the squared multiple correlation) values after entering the covariates range from .60 to .79 for the four cohorts, leaving very little reliable variance to be explained by the teacher variables. In each case it is the MAT premeasure that captures the largest share of background covariation with outcome, possibly inappropriately reducing any teacher effects that may be present: The pretest was administered

Table D.1

REGRESSIONS OF SPRING 1975 READING SCORES ON INDIVIDUAL AND AGGREGATE SPRING 1975 TEACHER SURVEY VARIABLES FOR TEACHERS AND STUDENTS IN ALTERNATIVE SCHOOLS

Variables[b] \ Grade Level[a]	Class Level				Program Level			
	3	4	5	6	3	4	5	6
Cohesion	-.026[c]	.068	.017	-.041	.091	.066	-.153*	.045
Principal Support	.000	-.044	-.089	.010	-.079	.027	-.054	.075
Autonomy	-.012	-.043	.052	-.007	.058	-.042	.049	.040
Task Oriented	-.159*[d]	-.037	-.009	-.052	-.049	.089	-.026	-.081
Principal Availability	.157	.059	-.043	-.014	.188	.070	-.064	.026
Pressure	.039	-.043	.108	.065	.073	-.057	-.005	-.009
Principal Control	.096	.068	-.026	-.003	.009	.092	-.080	-.010
Innovation	-.090	-.087	.050	.072	.093	.005	.006	.052
Demonstration Effect	.147	-.029	-.013	.036	-.079	-.093	.144	.000
Principal Influence	.129*	.031	-.081	-.005	.138	.060	-.176*	.037
Teacher Influence	-.152*	-.071	.000	-.001	-.112	-.076	-.057	-.050
Demo Problems	.026	.070	.087	.051	.050	.010	.074	.008
Common Policy	.014	-.023	-.050	-.000	-.035	-.045	.021	.004
N	280	282	320	362	383	393	443	450

[a]Grade level as of Spring 1976.

[b]See variable descriptions in Section II.

[c]Table entries are standardized in partial regression coefficients with covariate adjustment.

[d]Coefficients are asterisked if the corresponding unstandardized coefficients exceed twice their standard errors.

[e]Ns for analyses entering program aggregates are larger because a program score can be constructed for an individual child even if his/her teacher's scores are missing.

Table D.2

REGRESSIONS OF SPRING 1976 READING SCORES ON INDIVIDUAL AND AGGREGATE SPRING 1976 TEACHER SURVEY VARIABLES FOR TEACHERS AND STUDENTS IN ALTERNATIVE SCHOOLS

Variables \ Grade Level[a]	Class Level			Program Level		
	3	4	5	3	4	5
Cohesion	.005[c]	-.148	-.148*[d]	.125	-.054	-.117
Principal Support	-.057	-.045	-.027	.021	-.052	-.121
Autonomy	.067	.040	.020	.194	.200	-.013
Task Oriented	.083	.238*	.018	-.045	.045	.121
Principal Availability	-.072	-.164*	.031	.121	.088	.033
Pressure	-.083	.054	.063	-.085	-.029	-.013
Principal Control	-.059	-.107	-.048	-.023	-.099	.076
Innovation	-.057	.046	.055	-.277*	-.150*	-.014
Demonstration Effect	.028	-.016	-.066	.100	.180*	-.012
Principal Influence	-.008	-.082	-.033	.116	.010	-.046
Teacher Influence	-.023	.039	.054	-.313*	-.162*	.040
Demo Problems	.071	.127*	.014	-.030	.014	.049
Common Policy	-.058	-.116*	.056	.014	.024	-.009
N	388	328	405	482	453	506

[a]Grade level as of Spring 1976. The 6th grade cohort not included here both because students in regular schools transfer to alternative schools in 6th grade, and because many 6th grades shift to a high school format.

[b]Ns for analyses entering program aggregates are larger because a program score can be constructed for an individual child even if his/her teacher's scores are missing.

sometime during the fall of the same academic year as the outcome administration, and it is likely that its variation is, to some extent, a function of "current" teacher characteristics. Thus pretest-outcome covariation results partly from joint dependence on teacher variables, producing an adjustment that would bias our analysis toward null findings. Although this problem cannot be remedied for the analysis of third year effects (see the discussion in the Chapter II of data quality prior to the third year), the data permit avoiding it in the analyses to follow.

Table D.2 contains a set of analyses similar to those in the previous table, but for the spring 1976 teacher survey and student outcomes. In these analyses, MAT reading scores from the spring 1975 administration are used as premeasures, thereby avoiding inappropriate removal of outcome variation related to teacher effects. The oldest cohort of students is not represented in the table, since many sixth grade minischools offer high school style education, with no particular teacher associated with a given class.

Class and program level regressions of MAT reading outcomes from the spring of the fourth demonstration year on teacher survey responses obtained that same year are presented in Table D.2. Cohesiveness, Principal Support, and loose Principal Control appear to be associated with lower reading scores; however, only the first of these can be taken as a possible demonstration effect, since the latter two variables do not discriminate between participating and nonparticipating teachers (see Table 3.4). In contrast, Task Orientation and Autonomy, especially minischool-wide Autonomy, are related to positive reading outcomes.

Note that in no instance are the coefficients for the same variable significant at both levels of aggregation. Frequently here, a significant coefficient at one level is accomplished by an opposite signed or near zero-valued coefficient at the other level. Given that the program level coefficients are positively weighted components of the corresponding class level values, it is reasonable to conclude that the remaining portion of the class effect--the pooled within-program component--operates in an opposite direction to that of the between-program component.

Appendix E

TECHNICAL DETAILS OF THE ANALYSES

QUESTION 1: Are student outcomes different in regular and alternative schools?

A. Achievement outcomes

1. Analysis model: $Y = f(\underset{\sim}{X}, T)$

2. Variables and data sources:

Y--MAT total reading score, obtained in Spring 1975 (analysis interval Fall 1974-Spring 1975); Spring 1976 (analysis interval Fall 1974-Spring 1976); and Fall 1976 (analysis interval Fall 1974-Fall 1976).

$\underset{\sim}{X}^{*}$--(a) MAT total reading score, obtained in Fall 1974.
(b) Ethnicity: (i) Spanish surname
(ii) Spanish spoken at home
(c) SES: Aid to families with dependent children (AFDC).
(d) Attendance record: number of excused and inexcused days absent during analysis interval

T--Treatment group membership:

1 = Alternative (voucher) school
0 = Regular (non-voucher) school

3. Analysis method: Linear least squares regression treating individual student as the unit of analysis. Analyses were run separately for grade level cohorts. Ordinary standard errors from the (SPSS) regression program were used in determining significance levels.

4. Sample: All students enrolled in grades 3-6 in the fourth year of the demonstration with complete data on the variables used in the analysis.

B. Noncognitive outcomes

1. Analysis model: $Y = f(\underset{\sim}{X}, G, T, TG)$

2. Variables and data sources:

Y--Children's Self Social Constructs Test (CSSCT) (three scores defined in Chapter II); Spring 1976.

---

*Note: Four variants of $\underset{\sim}{X}$ were used as controls in the analyses: $X_{(a)}$ only; $X_{(a)}$ through $X_{(d)}$; and each of these correcting $X_{(a)}$ for errors of measurement (i.e., $TX_{(a)}$).

$\underset{\sim}{X}$--(a) CSSCT score corresponding to the particular Y used; Spring 1975.
(b)-(d)--same as 1A.
G--Grade level
T--Treatment group membership
TG--Treatment x grade level interaction (product)

3. Analysis method: Linear least squares regression treating individual student as unit of analysis. Ordinary standard errors from the regression program used for determining significance levels.

4. Sample: All in grades 3-6 with complete data.

QUESTION 2: What factors affect student achievement in alternative schools?

A. Teacher perceptions and attitudes

1. Analysis model: (a) $Y = f(\underset{\sim}{X}, \underset{\sim}{Z})$
(b) $Y = f(X, \overline{\underset{\sim}{Z}})$
(c) $\overline{Y} = f(\underset{\sim}{X}, G, \underset{\sim}{W}, \overline{\underset{\sim}{W}})$

2. Variables and data sources

Y--MAT total reading score for either Spring 1975 or Spring 1976 depending on the analysis.

$\underset{\sim}{X}$--(a)-(d) same as 1A.

$\overline{Y}$, $\overline{\underset{\sim}{X}}$--Class means on Y and $\underset{\sim}{X}$ (See p. 51 and footnote 7)

$\underset{\sim}{Z}$--Full set of teacher survey variables described in Chapter II.

$\overline{\underset{\sim}{Z}}$--Minischool means on $\underset{\sim}{Z}$

$\underset{\sim}{W}$--Selected subset of teacher survey variables

$\overline{\underset{\sim}{W}}$--Minischool means on $\underset{\sim}{W}$

G--Grade level (aggregate)

3. Analysis method: Models (a) and (b)--Linear least squares regression treating individual student as unit; analyses run separately for grade level cohorts. Model (c)--weighted least squares regression (see footnote b, Table 3.5 for weighting procedure) treating classrooms as analysis unit; analysis run for grade levels combined. All analyses use standard errors from the regression program for determining significance levels.

4. Sample: All students and teachers in alternative schools in grades 3-6 with complete data. For computation of $\bar{Z}$ and $\bar{W}$, all teachers in the minischool are used, whether or not their students appear in the analysis. Classrooms were excluded if fewer than 5 students had complete data; minischools were excluded having fewer than 2 classrooms with complete data.

B. Student perceptions of their classroom environment

1. Analysis model: $Y = f(\underset{\sim}{X}, \bar{\underset{\sim}{Z}})$

2. Variables and data sources:

Y--MAT total reading score; Spring 1976

$\underset{\sim}{X}$--(a)-(d) same as 1A.

$\bar{\underset{\sim}{Z}}$--Six scales from the Classroom Environment Survey (see Chapter II, pp. 24-25); aggregated to the classroom level.

3. Analysis method: Linear least squares regression treating individual student as analysis unit; analyses run separately for grade level cohorts. Significance levels from regular program.

4. Sample: All students with complete data in grades 3-6.

C. Program size

1. Analysis model: (a) $Y = f(\underset{\sim}{X}, S)$

(b) $Y = f(\underset{\sim}{X}, \underset{\sim}{Z}, S)$

2. Variables and data sources:

Y, $\underset{\sim}{X}$, $\underset{\sim}{Z}$--Same as Question 2A.

S--Minischool size: number of students enrolled in each minischool (only students with test records) in either the 3rd or 4th year of the demonstration; depending on the analysis.

3. Analysis method: Linear least squares regression treating individual student as analysis unit; analyses run separately for grade level cohorts. Significance levels determined from regression program standard errors.

4. Sample: Same as Question 2A.

D. Program choice options

1. Analysis model: (a) $Y = f(\underset{\sim}{X}, V, Z, Z_1, Z_2)$

(b) $Y = f(\underset{\sim}{X}, V, Z_1, Z_2, \bar{Z}_2, \bar{\bar{Z}}_2)$

2. Variables and data sources:

   Y, $\underset{\sim}{X}$--Same as Question 2B.

   V--Number of testing occasions prior to Fall 1974 each student was enrolled in an alternative school.

   Z_1--Depending on the analysis, either: (a) Number of times (testing occasions) student changed minischools, prior to Fall 1974;* or (b) Number of times student attended a minischool outside his/her catchment area, prior to Fall 1974.

   Z_2--Same as Z_1, (a) and (b), but for the interval Fall 1974 through Spring 1976.

   $\overline{Z}_2$--Class mean on Z_2.

   $\overline{\overline{Z}}_2$--Minischool mean on Z_2.

3. Analysis method: same as Question 2C.

4. Sample: All 3rd through 6th graders (as of Spring 1976) in alternative schools with complete data.

*Program change variables exclude changes between minischools in different catchment areas accompanied by corresponding family residence changes.

Table E.1

INDEX OF CROSS-REFERENCES BETWEEN TECHNICAL OUTLINE AND TEXT

Technical Outline	Text
Question 1	
A.	Table 3.1[a]
B.	Table 3.3
Question 2	
A.	Tables 3.5, D.1, D.2
(a)	Tables D.1 and D.2
(b)	Table 3.5
B.	Table 3.7
C.	Table 3.6
(a)	Column 2
(b)	Column 3
D.	Table 3.8
(a) Z_2	Row A.2
(b) Z_1	Row A.1
Z_2	Row B.1
$\overline{Z}_2$	Row B.2
$\overline{\overline{Z}}_2$	Row B.3

[a]See list of tables, pp. xvii-xviii, for page numbers.

REFERENCES

Alwin, D. F., "Structural Equation Models of Contextual Effects," paper presented at the Conference on Methodology for Aggregating Data in Educational Research, Stanford University, October 1976.

Alwin, D. F., and L.B. Otto, "High School Context Effects on Aspirations," Sociology of Education, Vol. 50, 1977, pp. 259-273.

Barker, P., T. K. Bikson, and J. Kimbrough, A Study of Alternatives in American Education, Vol. V: Diversity in the Classroom, The Rand Corporation, R-2170/5-NIE, February 1981.

Bass, G. V., A Study of Alternatives in American Education, Vol. I: District Policies and the Implementation of Change, The Rand Corporation, R-2170/1-NIE, April 1978.

Berman, P., and M. W. McLaughlin, Federal Programs Supporting Educational Change, Vol. VII: Factors Affecting Implementation and Continuation, The Rand Corporation, R-1589/7-HEW, April 1977.

Bikson, T. K., The Impact of Multiple Educational Options on Students' Self and Social Perceptions, The Rand Corporation, P-5860, April 1977.

Burstein, L., "The Role of Levels of Analysis in the Specifications of Educational Effects," in R. Dreeben and J. A. Thomas (eds.), Analysis of Educational Productivity, Vol. I: Issues in Microanalysis, Ballinger, Cambridge, Massachusetts, 1980.

Burstein, L., "Issues in the Aggregation of Data," in D. C. Berliner (ed.), Review of Research in Education, American Educational Research Association, Washington, D. C., 1980.

Cronbach, L. J., "Research on Classrooms and Schools: Formulation of Questions, Design, and Analysis," with the assistance of J. E. Deken and N. Webb, Occasional Paper, Stanford Evaluation Consortium, Stanford, California, July 1976.

Cronbach, L. J., D. Rogosa, R. E. Floden, and G. G. Price, "Analysis of Covariance in Non-Randomized Experiments: Parameters Affecting Bias," Occasional Paper, Stanford Evaluation Consortium, Stanford, California, August 1977.

Duncan, O. D., R. P. Cuzzort, and B. D. Duncan, Statistical Geography: Problems in Analyzing Areal Data, Free Press, Glencoe, 1961.

Dyer, H., R. L. Linn, and M. J. Patton, "A Comparison of Four Methods of Obtaining Discrepancy Measures Based on Observed and Predicted School System Means on Achievement Tests," American Educational Research Journal, Vol. 6, 1969, pp. 591-605.

Firebaugh, G. L., "Assessing Group Effects: A Comparison of Two Methods," Sociological Methods and Research, Vol. 7, 1979, pp. 384-395.

Hauser, R. M., Socioeconomic Background and Educational Performance, Rose Series Monographs, American Sociological Association, Washington, D. C., 1971.

Lincoln, J. R., and G. Zeitz, "Organizational Properties from Aggregate Data: Separating Individual and Structural Effects," paper presented at the Annual Meeting of the American Sociological Association, Boston, Massachusetts, 1979.

Lord, F. M., "A Paradox in the Interpretation of Group Comparisons," Psychological Bulletin, Vol. 68, 1967, pp. 304-305.

Moos, R., and P. Insel, The Work Environment Scale, Social Ecology Laboratory, Department of Psychiatry, Stanford University, Palo Alto, 1974.

Rasmussen, R. L., A Study of Alternatives in American Education, Vol. III: Teachers' Response to Alternatives, The Rand Corporation, R-2170/3-NIE, February 1981.

Snow, R. E., "Representative and Quasi-Representative Designs for Research in Teaching," Review of Educational Research, Vol. 44, 1974, pp. 265-291.

Wortman, P. M., and R. St. Pierre, "The Education Voucher Demonstration: A Secondary Analysis," Education and Urban Society, Vol. 9, 1977, pp. 471-491.

Ziller, R. C., The Social Self, Pergamon Press, New York, 1973.